EXHIBIT A

US011470243B2

(12) **United States Patent**
Nielsen

(10) **Patent No.: US 11,470,243 B2**
(45) **Date of Patent: Oct. 11, 2022**

(54) **METHODS AND APPARATUS TO CAPTURE IMAGES**

(71) Applicant: **The Nielsen Company (US), LLC**, New York, NY (US)

(72) Inventor: **Christen V. Nielsen**, Palm Harbor, FL (US)

(73) Assignee: **The Nielsen Company (US), LLC**, New York, NY (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **17/751,283**

(22) Filed: **May 23, 2022**

(65) **Prior Publication Data**

US 2022/0286601 A1 Sep. 8, 2022

Related U.S. Application Data

(63) Continuation of application No. 17/666,322, filed on Feb. 7, 2022, which is a continuation of application

(Continued)

(51) **Int. Cl.**

H04N 5/232	(2006.01)
G06V 40/16	(2022.01)
H04N 21/4223	(2011.01)
H04N 21/442	(2011.01)
G06V 40/10	(2022.01)
G06V 20/52	(2022.01)
H04N 21/4415	(2011.01)
H04H 60/45	(2008.01)
G06V 10/141	(2022.01)
H04N 5/347	(2011.01)

(Continued)

(52) **U.S. Cl.**

CPC ***H04N 5/23219*** (2013.01); ***G06V 10/141*** (2022.01); ***G06V 20/53*** (2022.01); ***G06V 40/103*** (2022.01); ***G06V 40/161*** (2022.01); ***G06V 40/172*** (2022.01); ***H04H 60/45*** (2013.01); ***H04N 5/2256*** (2013.01); ***H04N 5/33*** (2013.01); ***H04N 5/347*** (2013.01); ***H04N 21/4223*** (2013.01); ***H04N 21/4415*** (2013.01); ***H04N 21/44218*** (2013.01)

(58) **Field of Classification Search**

CPC H04N 5/23219; H04N 5/2256; H04N 5/33; H04N 5/347; H04N 21/4223; H04N 21/4415; H04N 21/44218; G06V 10/141; G06V 20/53; G06V 40/103; G06V 40/161; G06V 40/172; H04H 60/45

See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

5,987,154 A 11/1999 Gibbon et al.
6,014,461 A 1/2000 Hennessey et al.

(Continued)

OTHER PUBLICATIONS

Wikipedia. (Dec. 2010) "Kinect." http://en.wikipedia.org/wiki/Kinect, 15 pages.

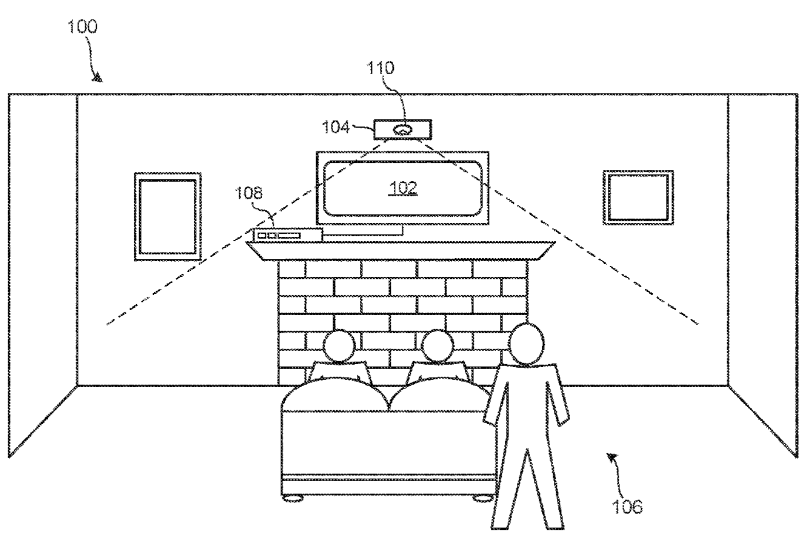
(Continued)

Primary Examiner — Amandeep Saini
(74) *Attorney, Agent, or Firm* — Hanley, Flight & Zimmerman, LLC

(57) **ABSTRACT**

Methods and apparatus to obtain exposure data for media exposure environment(s) are disclosed. An example apparatus disclosed herein includes processor circuitry to determine audience identification information and content identifying data.

29 Claims, 5 Drawing Sheets



Related U.S. Application Data

No. 16/878,935, filed on May 20, 2020, now Pat. No. 11,245,839, which is a continuation of application No. 16/196,810, filed on Nov. 20, 2018, now abandoned, which is a continuation of application No. 15/793,108, filed on Oct. 25, 2017, now Pat. No. 10,165,177, which is a continuation of application No. 15/419,120, filed on Jan. 30, 2017, now Pat. No. 9,843,717, which is a continuation of application No. 14/732,175, filed on Jun. 5, 2015, now Pat. No. 9,560,267, which is a continuation of application No. 13/327,227, filed on Dec. 15, 2011, now Pat. No. 9,082,004.

(51) **Int. Cl.**

H04N 5/33 (2006.01)
H04N 5/225 (2006.01)

(56) **References Cited**

U.S. PATENT DOCUMENTS

7,043,056 B2	5/2006	Edwards et al.	
7,440,593 B1	10/2008	Steinberg et al.	
7,602,524 B2	10/2009	Eichhorn et al.	
7,636,456 B2	12/2009	Collins et al.	
7,676,065 B2	3/2010	Wiedemann et al.	
7,697,735 B2	4/2010	Adam et al.	
7,796,154 B2	9/2010	Senior et al.	
7,899,209 B2	3/2011	Greiffenhagen et al.	
8,620,088 B2	12/2013	Lee	
9,082,004 B2	7/2015	Nielsen	
9,560,267 B2	1/2017	Nielsen	
9,843,717 B2	12/2017	Nielsen	
10,165,177 B2	12/2018	Nielsen	
11,245,839 B2	2/2022	Nielsen	
2003/0081834 A1	5/2003	Philomin et al.	
2003/0093784 A1	5/2003	Dimitrova et al.	
2004/0122679 A1	6/2004	Neuhauser et al.	
2005/0117783 A1	6/2005	Sung et al.	
2006/0062429 A1	3/2006	Ramaswamy et al.	
2006/0133699 A1	6/2006	Widrow et al.	
2007/0150916 A1	6/2007	Begole et al.	
2007/0154063 A1	7/2007	Breed	
2007/0263934 A1	11/2007	Ojima et al.	
2007/0294126 A1	12/2007	Maggio	
2008/0232650 A1	9/2008	Suzuki et al.	
2008/0243590 A1	10/2008	Rich	
2008/0271065 A1	10/2008	Buonasera et al.	
2009/0091650 A1	4/2009	Kodama	
2009/0177528 A1	7/2009	Wu et al.	
2009/0307084 A1	12/2009	Monighetti et al.	
2009/0310829 A1	12/2009	Baba et al.	
2010/0245567 A1	9/2010	Kralmstoever et al.	
2011/0019924 A1	1/2011	Elgersma et al.	
2011/0137721 A1	6/2011	Bansal	
2011/0164188 A1*	7/2011	Karaoguz	H04N 21/4122 348/734
2011/0169953 A1	7/2011	Sandler et al.	
2011/0265110 A1	10/2011	Weinblatt	
2011/0285845 A1	11/2011	Bedros et al.	
2012/0081392 A1	4/2012	Arthur	
2012/0151079 A1	6/2012	Besehanic et al.	
2013/0013396 A1	1/2013	Vinson et al.	
2013/0129159 A1	5/2013	Huijgens et al.	
2013/0152113 A1*	6/2013	Conrad	H04N 21/42201 725/12
2013/0156273 A1	6/2013	Nielsen	
2015/0271390 A1	9/2015	Nielsen	
2017/0142330 A1	5/2017	Nielsen	
2018/0048807 A1	2/2018	Nielsen	
2019/0089894 A1	3/2019	Nielsen	
2020/0351436 A1	11/2020	Nielsen	

OTHER PUBLICATIONS

Fellers et al. (Aug. 2004) "Nikon MicroscopyU: Interactive Java Tutorials—CCD Signal-to-Noise Ratio." http://www.microscopyu.com/tutorials/java/digitalimaging/signaltonoise/index.html as archived by The Internet Archive, www.archive.org., 4 pages.

United States Patent and Trademark Office, "Non-Final Office Action," issued in connection with U.S. Appl. No. 13/327,227, dated May 20, 2014, 15 pages.

United States Patent and Trademark Office, "Final Office Action," issued in connection with U.S. Appl. No. 13/327,227, dated Sep. 18, 2014, 20 pages.

United States Patent and Trademark Office, "Notice of Allowance and Fee(s) Due," issued in connection with U.S. Appl. No. 13/327,227, dated Mar. 6, 2015, 9 pages.

United States Patent and Trademark Office, "Requirement for Restriction/Election," issued in connection with U.S. Appl. No. 14/732,175, dated Jul. 8, 2016, 6 pages.

United States Patent and Trademark Office, "Notice of Allowance and Fee(s) Due," issued in connection with U.S. Appl. No. 14/732,175, dated Sep. 23, 2016, 10 pages.

United States Patent and Trademark Office, "Notice of Allowance and Fee(s) Due," issued in connection with U.S. Appl. No. 15/419,120, dated Aug. 2, 2017, 10 pages.

United States Patent and Trademark Office, "Notice of Allowance and Fee(s) Due," issued in connection with U.S. Appl. No. 15/793,108, dated Aug. 23, 2018, 10 pages.

United States Patent and Trademark Office, "Non-Final Office action," issued in connection with U.S. Appl. No. 15/793,108, dated Mar. 2, 2018, 11 pages.

United States Patent and Trademark Office, "Advisory action," issued in connection with U.S. Appl. No. 16/196,810, dated May 13, 2020, 2 pages.

United States Patent and Trademark Office, "Final Office action," issued in connection with U.S. Appl. No. 16/196,810, dated Feb. 26, 2020, 23 pages.

United States Patent and Trademark Office, "Non-Final Office action," issued in connection with U.S. Appl. No. 16/196,810, dated Nov. 7, 2019, 15 pages.

United States Patent and Trademark Office, "Advisory action," issued in connection with U.S. Appl. No. 16/196,810, dated Aug. 22, 2019, 2 pages.

United States Patent and Trademark Office, "Final Office action," issued in connection with U.S. Appl. No. 16/196,810, dated Jun. 13, 2019, 16 pages.

United States Patent and Trademark Office, "Non-Final Office action," issued in connection with U.S. Appl. No. 16/196,810, dated Mar. 1, 2019, 14 pages.

United States Patent and Trademark Office, "Non-Final Office action," issued in connection with U.S. Appl. No. 16/878,935, dated Jun. 10, 2021, 9 pages.

United States Patent and Trademark Office, "Notice of Allowance and Fee(s) Due," issued in connection with U.S. Appl. No. 16/878,935, dated Sep. 27, 2021, 7 pages.

* cited by examiner

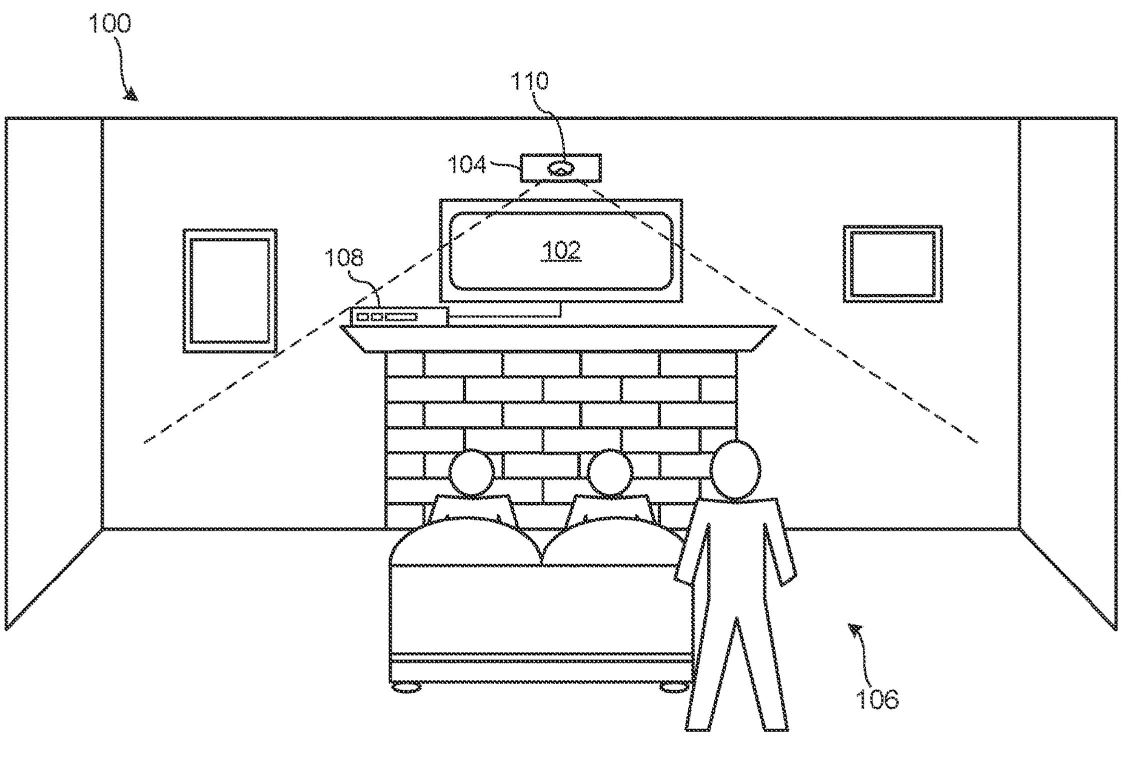


FIG. 1

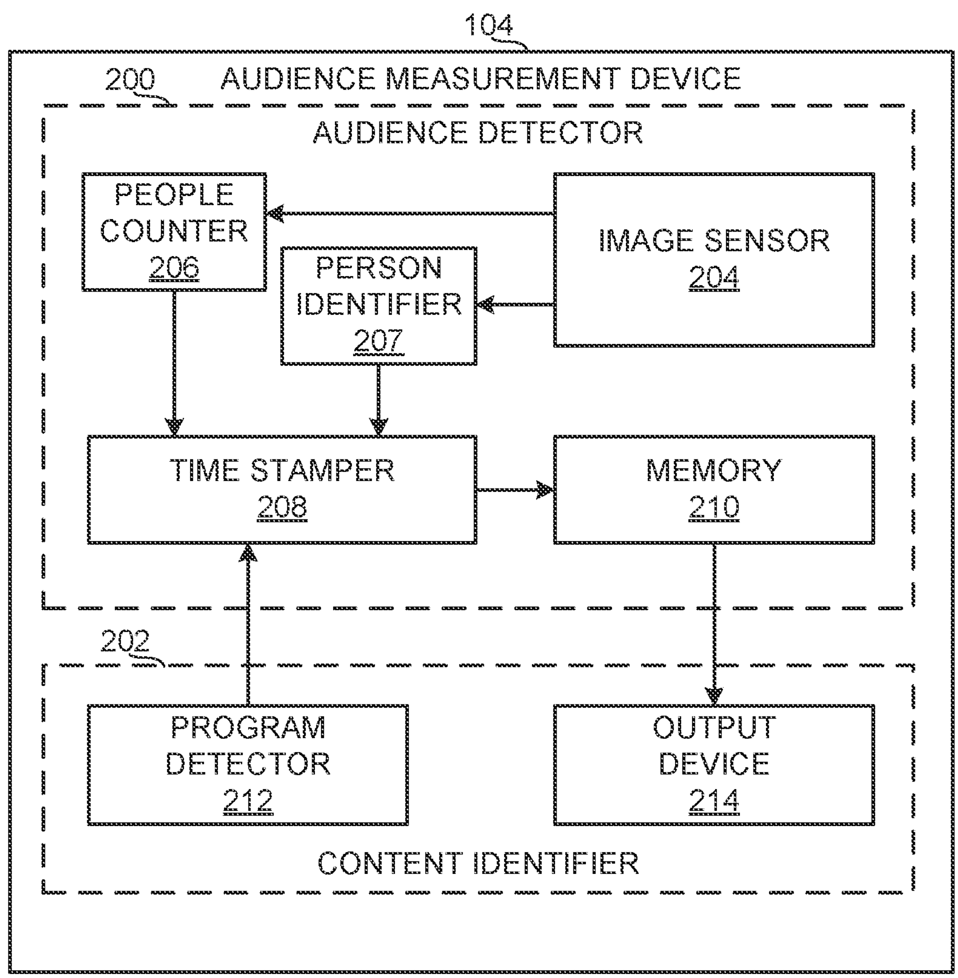

104
200
AUDIENCE MEASUREMENT DEVICE
AUDIENCE DETECTOR
PEOPLE COUNTER 206
PERSON IDENTIFIER 207
IMAGE SENSOR 204
TIME STAMPER 208
MEMORY 210
202
PROGRAM DETECTOR 212
OUTPUT DEVICE 214
CONTENT IDENTIFIER

FIG. 2

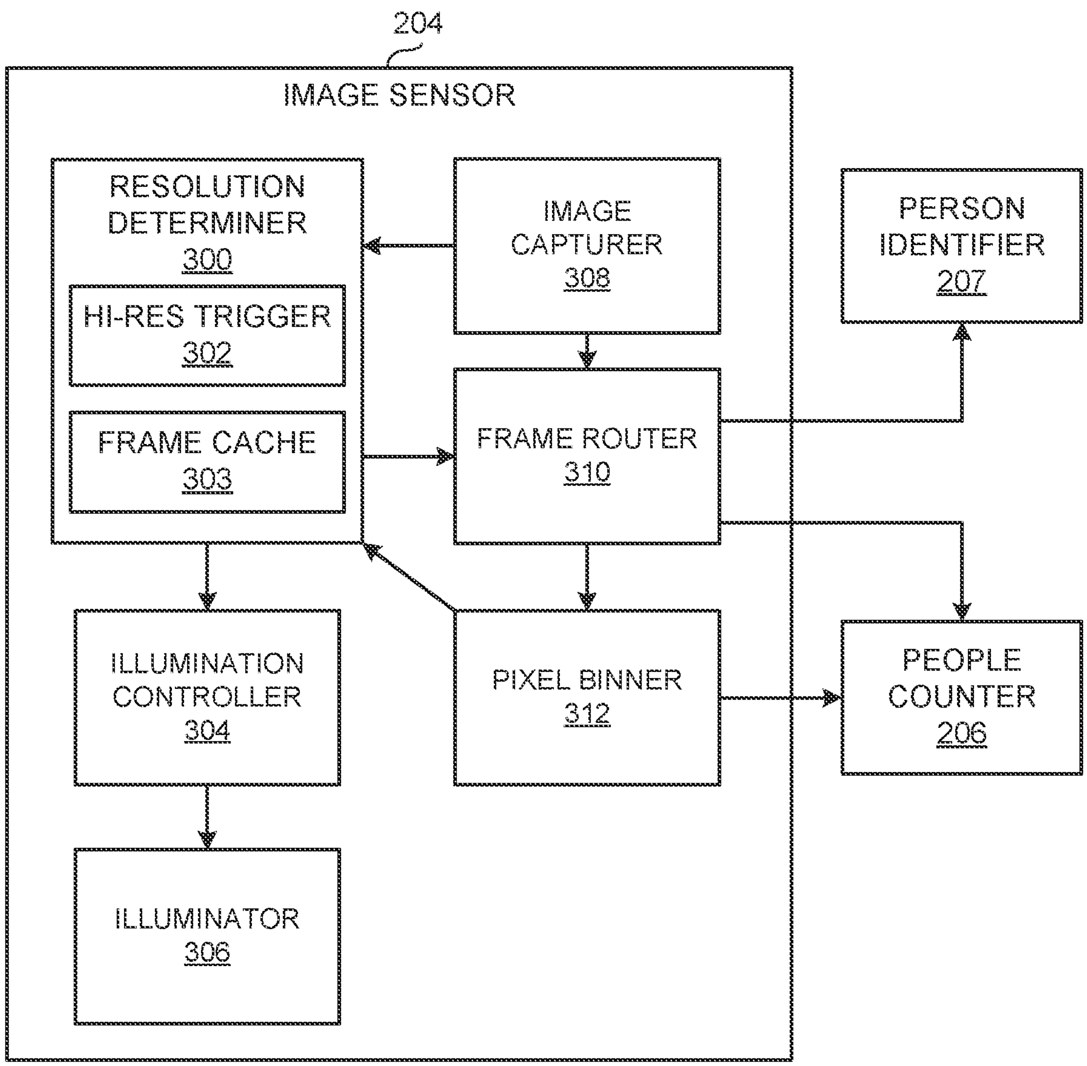
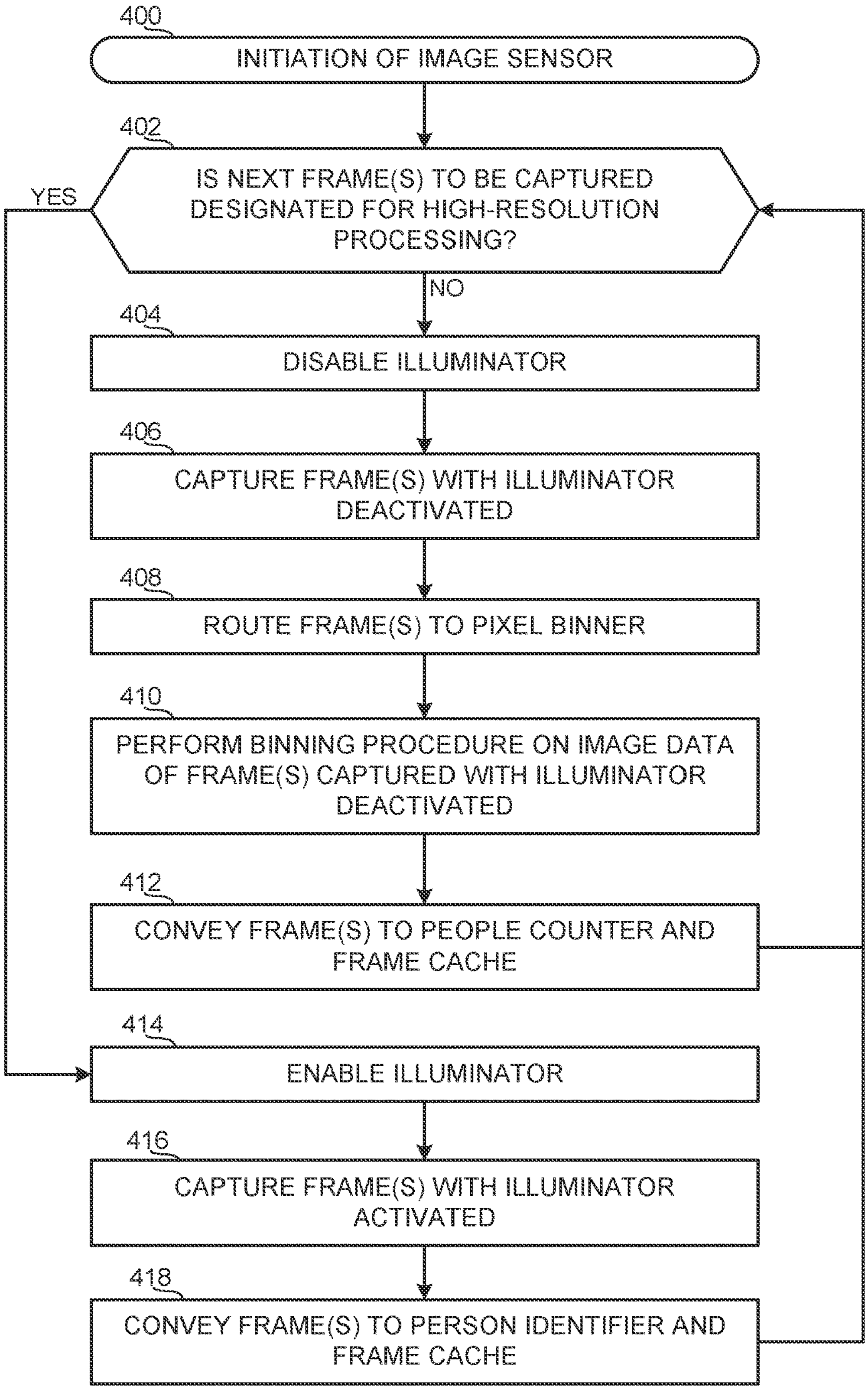

204
IMAGE SENSOR
RESOLUTION DETERMINER 300
HI-RES TRIGGER 302
FRAME CACHE 303
IMAGE CAPTURER 308
FRAME ROUTER 310
PERSON IDENTIFIER 207
ILLUMINATION CONTROLLER 304
PIXEL BINNER 312
PEOPLE COUNTER 206
ILLUMINATOR 306

FIG. 3



400
INITIATION OF IMAGE SENSOR
402
IS NEXT FRAME(S) TO BE CAPTURED DESIGNATED FOR HIGH-RESOLUTION PROCESSING?
YES
NO
404
DISABLE ILLUMINATOR
406
CAPTURE FRAME(S) WITH ILLUMINATOR DEACTIVATED
408
ROUTE FRAME(S) TO PIXEL BINNER
410
PERFORM BINNING PROCEDURE ON IMAGE DATA OF FRAME(S) CAPTURED WITH ILLUMINATOR DEACTIVATED
412
CONVEY FRAME(S) TO PEOPLE COUNTER AND FRAME CACHE
414
ENABLE ILLUMINATOR
416
CAPTURE FRAME(S) WITH ILLUMINATOR ACTIVATED
418
CONVEY FRAME(S) TO PERSON IDENTIFIER AND FRAME CACHE

FIG. 4

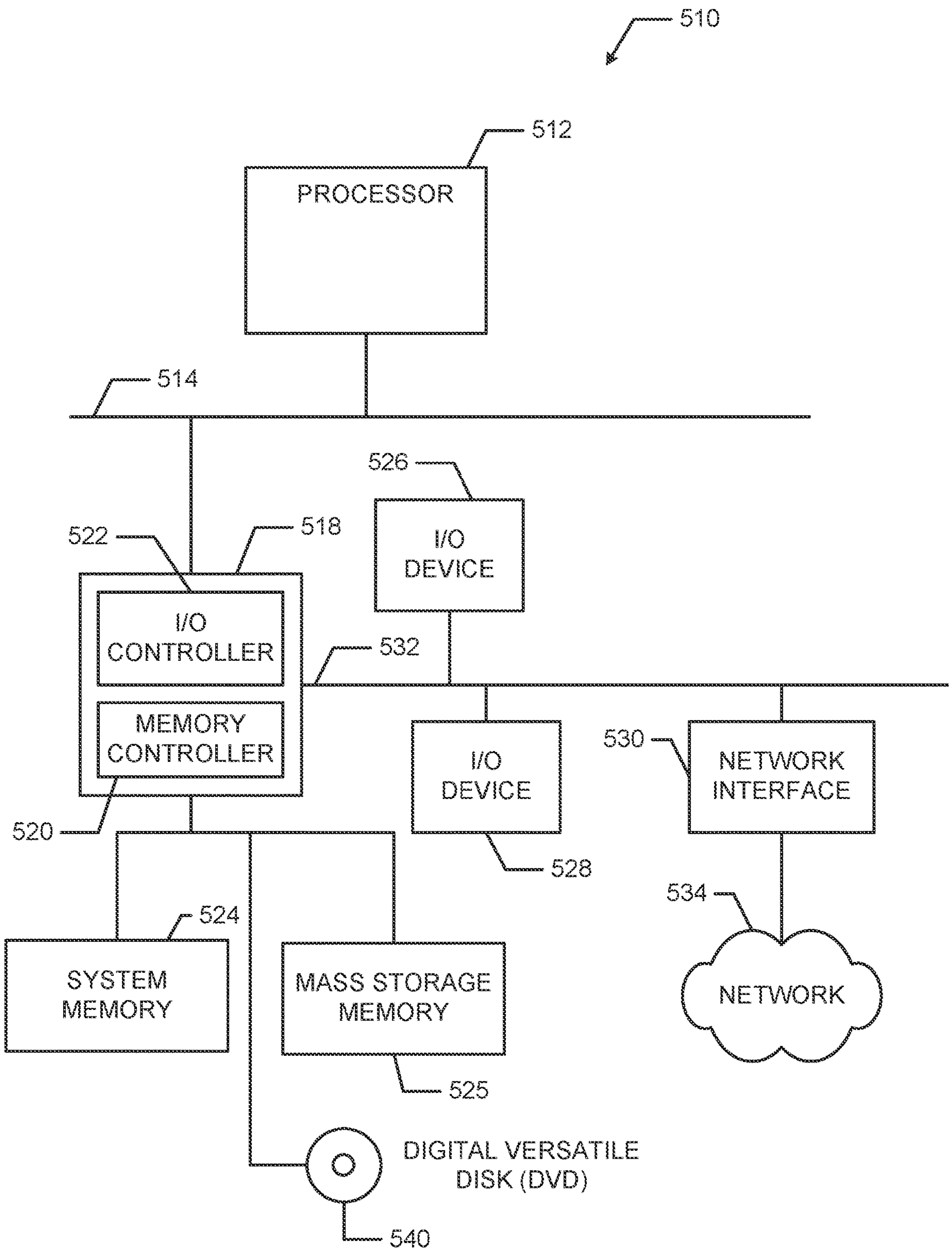

510
512
PROCESSOR
514
526
I/O
DEVICE
522
518
I/O
CONTROLLER
532
MEMORY
CONTROLLER
I/O
DEVICE
530
NETWORK
INTERFACE
520
528
534
524
SYSTEM
MEMORY
MASS STORAGE
MEMORY
NETWORK
525
DIGITAL VERSATILE
DISK (DVD)
540

FIG. 5

METHODS AND APPARATUS TO CAPTURE IMAGES

RELATED APPLICATIONS

This patent arises from a continuation of U.S. patent application Ser. No. 17/666,322, filed Feb. 7, 2022, which is a continuation of U.S. patent application Ser. No. 16/878,935, filed May 20, 2020, now U.S. Pat. No. 11,245,839, which is a continuation of U.S. patent application Ser. No. 16/196,810, filed Nov. 20, 2018, which is a continuation of U.S. patent application Ser. No. 15/793,108, filed Oct. 25, 2017, now U.S. Pat. No. 10,165,177, which is a continuation of U.S. patent application Ser. No. 15/419,120, filed Jan. 30, 2017, now U.S. Pat. No. 9,843,717, which is a continuation of U.S. patent application Ser. No. 14/732,175, filed Jun. 5, 2015, now U.S. Pat. No. 9,560,267, which is a continuation of U.S. patent application Ser. No. 13/327,227, filed Dec. 15, 2011, now U.S. Pat. No. 9,082,004. U.S. patent application Ser. No. 17/666,322, U.S. patent application Ser. No. 16/878,935, U.S. patent application Ser. No. 16/196,810, U.S. patent application Ser. No. 15/793,108, U.S. patent application Ser. No. 15/419,120, U.S. patent application Ser. No. 14/732,175, and U.S. patent application Ser. No. 13/327,227 are hereby incorporated herein by reference in their entirety. Priority to U.S. patent application Ser. No. 17/666,322, U.S. patent application Ser. No. 16/878,935, U.S. patent application Ser. No. 16/196,810, U.S. patent application Ser. No. 15/793,108, U.S. patent application Ser. No. 15/419,120, U.S. patent application Ser. No. 14/732,175 and U.S. patent application Ser. No. 13/327,227 is claimed.

FIELD OF THE DISCLOSURE

This disclosure relates generally to audience measurement and, more particularly, to methods and apparatus to capture images.

BACKGROUND

Audience measurement of media content (e.g., broadcast television and/or radio, stored audio and/or video content played back from a memory such as a digital video recorder or a digital video disc, audio and/or video content played via the Internet, video games, etc.) often involves collection of content identifying data (e.g., signature(s), fingerprint(s), embedded code(s), channel information, time of consumption information, etc.) and people data (e.g., identifiers, demographic data associated with audience members, etc.). The content identifying data and the people data can be combined to generate, for example, media exposure data indicative of amount(s) and/or type(s) of people that were exposed to specific piece(s) of media content.

In some audience measurement systems, the collected people data includes an amount of people being exposed to media content. To calculate the amount of people being exposed to the media content, some measurement systems capture a series of images of a media exposure environment (e.g., a television room, a family room, a living room, a bar, a restaurant, etc.) and analyze the images to determine how many people appear in the images at a particular date and time. The calculated amount of people in the media exposure environment can be correlated with media content being presented at the particular date and time to provide exposure data (e.g., ratings data) for that media content. Additionally, some audience measurement systems identify people in the images via one or more identification techniques such as, for example, facial recognition.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1** is an illustration of an example exposure environment including an example audience measurement device disclosed herein.

FIG. **2** is a block diagram of an example implementation of the example audience measurement device of FIG. **1**.

FIG. **3** is a block diagram of an example implementation of the example image sensor of FIG. **2**.

FIG. **4** is a flowchart illustrating example machine readable instructions that may be executed to implement the example image sensor of FIGS. **1**, **2** and/or **3**.

FIG. **5** is a block diagram of an example processing system capable of executing the example machine readable instructions of FIG. **4** to implement the example image sensor of FIGS. **1**, **2** and/or **3**.

DETAILED DESCRIPTION

To count people in a media exposure environment, such as a room of a house in which a television is located, some audience measurement systems attempt to recognize objects as humans in a series of captured images of the room. A tally is maintained for each frame of image data to reflect an amount of people in the room at a time corresponding to a respective frame. That is, each recognition of an object as a human in a frame increases the tally associated with that frame. The audience measurement system counting the people in the room may also collect content identifying information to identify media content being presented (e.g., aurally and/or visually) in the room. With the identification of the media content and the amount of people in the room at a given date and time, the audience measurement system is aware of how many people were exposed to the specific media content.

Additionally, some systems recognize an identity of one or more of the detected humans by, for example, performing a facial recognition process on image data of one or more of the frames, receiving identifier(s) from the detected humans, detecting identifying signal(s) generated by devices carried by the humans, etc. Personal identification information can be used in conjunction with the content identifying information and/or the tally information to generate exposure information related to the content. When an audience measurement system uses a facial recognition process to identify people, an accuracy of the identification increases with an increase in resolution of the image data on which the facial recognition process is performed. In other words, the higher the resolution of a frame of image data, the more likely identification made via facial recognition will be accurate.

To provide high-resolution image data, audience measurement systems that include facial recognition capabilities typically employ high-resolution image sensors equipped with an illumination source, such as an infrared (IR) light emitting diode (LED). In previous systems, each time the high-resolution image sensor captures a frame of image data, the illumination source illuminates a surrounding area. The resulting illumination provides lighting conditions favorable to capturing high-resolution image data. When the illumination source is an IR LED, the illumination source emits IR light to enable the image sensor to capture illuminated

objects. In addition to the IR light, IR emitters also emit light from the visible spectrum that appears as a red glow from the emitter.

Frequent activation of the illumination sources when capturing image data represents a significant power drain for the audience measurement system. Moreover, frequent interval activation (e.g., every two seconds) of the illumination source shortens a lifetime of the illumination source. Additionally, due to significant amounts of heat generated by the illumination sources, heat sinking devices and techniques are typically needed in systems that activate the illumination sources frequently. Further, light emissions of the illumination source has the potential to annoy people such as, for example, members of a panel that are exposed to the illumination source while in the presence of the audience measurement system. In an audience measurement system utilizing an IR LED, the red glow emitted from the illumination source is a blinking red light that faces the panel members while the panel members are, for example, watching a television. This blinking red light may annoy some panel members. Annoyance of panel members is undesirable and may prove detrimental to an ability of the audience measurement system to maintain persons in the panel and/or to collect as much data as possible. That is, some audience measurement systems rely on the willing participation of panel members and, thus, reduction or elimination of annoying aspects of the system is beneficial to avoid impairing willingness to volunteer. An annoying feature of the audience measurement system may decrease panelist compliance with and/or participation in the system.

Example methods, apparatus, and articles of manufacture disclosed herein reserve use of an illumination source of a high-resolution image sensor for frames of image data designated for processing that requires high-resolution image data. Example frames designated for such processing include frames on which a facial recognition process is to be executed. For frames not designated for processing that requires high-resolution image data, example methods, apparatus, and articles of manufacture disclosed herein capture image data without the use of the illumination source. Example frames not designated for processing that requires high-resolution image data include frames on which a person count (e.g., a body count) is to be executed without recognizing an identity of the detected persons. Additionally, when image data is captured without activation of the illumination source (e.g., when the corresponding frame will not be subjected to facial recognition), example methods, apparatus, and articles of manufacture disclosed herein enhance resulting images to compensate for low light levels and loss of contrast due to lack of illumination. In particular, example methods, apparatus, and articles of manufacture disclosed herein employ a pixel binning procedure on the image data captured without use of the illumination source. Binning is the process of summing pixel values in a neighborhood of pixels (e.g., a 2×2, 3×3, 4×4, etc. area), thereby capturing more light per pixel at the cost of lower resolution. However, for example methods, apparatus, and articles of manufacture disclosed herein, the lower resolution is acceptable for the frames captured without the use of the illumination source because, as described above, those frames will not be subjected to processing that requires high-resolution data.

Accordingly, example methods, apparatus, and articles of manufacture disclosed herein selectively activate an illumination source associated with an image sensor for certain frames. In contrast, previous systems activate the illumination source for each frame captured by a corresponding

image sensor. In many instances, frames not designated for high-resolution processing (e.g., frames used solely to count people) greatly outnumber the frames designated for high-resolution frame (e.g., frame used for facial recognition). In fact, as described below, a mode of an image sensor during which image data is captured without the use of the illumination source (e.g., because high-resolution image data is not necessary for the corresponding frame(s)) is referred to herein as a ‘majority capture’ mode. Conversely, because the number of frames requiring high-resolution image data is far less than the number of frames for which lower resolution image data is acceptable, a mode of the image sensor during which image data is captured with the use of the illumination source is referred to herein as a ‘minority capture’ mode.

Because the illumination source is a significant power consumer, selective activation of the illumination source provided by example methods, apparatus, and articles of manufacture disclosed herein greatly reduce power consumption levels of the image sensors of audience measurement systems. Moreover, the selective activation provided by example methods, apparatus, and articles of manufacture disclosed herein extends the operational lifetime of the image sensors of audience measurement systems by less frequently operating the corresponding illumination sources. Further, the selective activation provided by example methods, apparatus, and articles of manufacture disclosed herein reduces or even eliminates the need for heat sinking devices and/or techniques otherwise required to dissipate heat generated by the illumination sources. In addition to the reduced resource consumption provided by example methods, apparatus, and articles of manufacture disclosed herein, audience measurement methods, systems, and articles of manufacture employing the selective activation disclosed herein reduce the likelihood that panel members become irritated by light and/or glow emitted from illumination sources. As described above, illumination sources of image sensors typically face panel members in a media exposure environment (e.g., a television room) and, thus, the panel members are subjected to a blinking and/or glowing light whenever the illumination source is activated. In previous systems utilizing a high-resolution image sensor, each frame is captured with the use of the illumination source, resulting in a blinking light (e.g., red light in the case of an IR LED flash unit being used as the illumination source) or a light that is seemingly persistently on throughout operation of the system. Selectively activating the illumination source for frames requiring high-resolution image data and not activating the illumination source for other frames, as disclosed herein, considerably reduces the instances of illumination of the media exposure environment and, thus, the potential for irritating the panel members.

FIG. **1** is an illustration of an example media exposure environment **100** including a media presentation device **102** and an example audience measurement device **104** for measuring and/or identifying an audience **106** of the media presentation device **102**. In the illustrated example of FIG. **1**, the media exposure environment **100** is a room of a household that has been statistically selected to develop television ratings data for a population/demographic of interest. Assumingly, one or more persons of the household have registered with the system and provided this demographic information. The example audience measurement device **104** can be implemented in additional and/or alternative types of environments such as, for example, a room in a non-statistically selected household, a theater, a restaurant, a tavern, a retail location, an arena, etc. In the illustrated example of FIG. **1**, the media presentation device is a

television **102** coupled to a set-top box (STB) **108** that implements a digital video recorder (DVR) and a digital versatile disc (DVD) player. The example audience measurement device **104** can be implemented in connection with additional and/or alternative types of media presentation devices such as, for example, a radio, a computer monitor, a video game console and/or any other communication device able to present content to one or more individuals.

The example audience measurement device **104** of FIG. **1** utilizes a camera **110** to capture a plurality of time stamped frames of image data of the environment **100**. The example camera **110** captures images within a field of view defined by the dotted lines of FIG. **1**. In the example shown in FIG. **1**, an image captured by the camera **110** includes each member of a three-person audience **106**. The images captured by the camera **110** are used to generate people tallies representative of how many people are in the audience **106** and/or personal identifications of the people in the audience. As described in detail below, the example audience measurement device **104** of FIG. **1** also monitors the environment **100** to identify media content being presented (e.g., displayed, played, etc.) by the television **102** and/or other media presentation devices to which the audience **106** is exposed. Identification(s) of media content to which the audience **106** is exposed are correlated with the people tallies and/or the personal identifications to generate exposure data for the media content.

FIG. **2** is a block diagram of an example implementation of the example audience measurement device **104** of FIG. **1**. The example audience measurement device **104** of FIG. **2** includes an audience detector **200** and a content identifier **202**. The example audience detector **200** includes an image sensor **204**, a people counter **206**, a person identifier **207**, a time stamper **208**, and a memory **210**. The example image sensor **204** of FIG. **2** captures frames of image data of the environment **100**, which includes the audience **106** being exposed to a presentation output by the media presentation device **102** of FIG. **1**. In the illustrated example, the image sensor **204** is implemented by a high-resolution camera capable of capturing high-resolution image data, such as an image sensor configured to capture images at a resolution of, for example, 1920×1080 or 1280×960. In some examples, the image sensor **204** and/or the camera **110** of FIG. **1** is implemented by a gaming system, such as XBOX® Kinect®.

In the illustrated example, the frames obtained by the image sensor **204** of FIG. **2** are conveyed to the people counter **206**. The example people counter **206** determines how many people appear in each of the received frames and records each of the amounts of people as a tally for each frame. The example people counter **206** can determine how many people appear in a frame in any suitable manner using any suitable technique. For example, the people counter **206** of FIG. **2** recognizes a general shape of a human body and/or a human body part, such as a head and/or torso. Additionally or alternatively, the example people counter **206** of FIG. **2** may count a number of "blobs" that appear in the frame and count each distinct blob as a person. Recognizing human shapes and counting "blobs" are illustrative examples and the people counter **206** of FIG. **2** can count people using any number of additional and/or alternative techniques. An example manner of counting people is described by Blumenthal in U.S. patent application No. 10/538,483, filed on Jun. 8, 2005, now U.S. Pat. No. 7,203,338, which is hereby incorporated herein by reference in its entirety. To track the number of detected people in a room, the example people counter **206** of FIG. **2** also tracks a position (e.g., an X-Y coordinate) of each detected person.

In the illustrated example, some frames obtained by the image sensor **204** of FIG. **2** are conveyed to the person identifier **207** in addition to or in lieu of the people counter **206**. As described in greater detail below in connection with FIG. **3**, the frames conveyed to the example person identifier **207** of FIG. **2** are frames designated (e.g., by the image sensor **204** and/or any other component of the audience measurement system **100**) for a facial recognition procedure such that people captured in those frames can be individually identified. In some examples, the audience detector **200** may have additional or alternative methods and/or components to identify people in the frames captured by the image sensor **204**, such as a feedback system to which the members of the audience **106** provide (e.g., actively and/or passively) identification to the audience measurement device **104**). To identify people in the frames captured by the image sensor **204**, the example person identifier **207** includes or has access to a collection (e.g., stored in a database) of facial signatures (e.g., vectors). Each facial signature corresponds to a person having a known identity to the person identifier **207**. The collection includes an identifier (ID) for each known facial signature that corresponds to a known person. For example, in reference to FIG. **1**, the collection of facial signatures may correspond to frequent visitors and/or members of the household associated with the room **100**. The example person identifier **207** analyzes one or more regions of a frame thought to correspond to a human face and develops a pattern or map for the region(s). The pattern or map of the region represents a facial signature of the detected human face. The example person identifier **207** compares the detected facial signature to entries of the facial signature collection. When a match is found, the person identifier **207** has successfully identified at least one person in the frame. In such instances, the person identifier **207** records (e.g., in a memory address dedicated to the person identifier **207**) the ID associated with the matching facial signature of the collection. When a match is not found, the person identifier **207** of the illustrated example retries the comparison or prompts the audience **106** for information that can be added to the collection of known facial signatures for the unmatched face. More than one signature may correspond to the same face (i.e., the face of the same person).

The example people counter **206** of FIG. **2** outputs the calculated tallies and/or corresponding image frames (or identifications thereof) to the time stamper **208**. Also, the person identifier **207** outputs the records ID(s) for any identified persons and/or the corresponding image frames (or identifications thereof) to the time stamper **208**. The time stamper **208** of the illustrated example includes a clock and a calendar. The example time stamper **208** associates a time and date with each calculated tally and/or ID and the corresponding frame by, for example, appending the time/date data to the end of the tally data, the ID(s), and/or the frame. A data package (e.g., the tally, the ID(s), the date and time, and/or the frame) is stored in the memory **210**. The memory **210** may include a volatile memory (e.g., Synchronous Dynamic Random Access Memory (SDRAM), Dynamic Random Access Memory (DRAM), RAMBUS Dynamic Random Access Memory (RDRAM, etc.) and/or a non-volatile memory (e.g., flash memory). The memory **210** may include one or more double data rate (DDR) memories, such as DDR, DDR2, DDR3, mobile DDR (mDDR), etc. The memory **210** may also include one or more mass storage devices such as, for example, hard drive disk(s), compact disk drive(s), digital versatile disk drive(s), etc.

The example content identifier **202** of FIG. **2** includes a program detector **212** and an output device **214**. The example program detector **212** of FIG. **2** detects presentation(s) of media content in the media exposure environment **100** and collects identification information associated with the detected presentation(s). For example, the program detector **212**, which may be in wired and/or wireless communication with the presentation device **102** and/or the STB **108** of FIG. **1**, can identify a presentation time and a source of a presentation. The presentation time and the source identification data may be utilized to identify the program by, for example, cross-referencing a program guide configured, for example, as a look up table. The source identification data may, for example, be the identity of a channel obtained, for example, by monitoring a tuner of the STB **108** or a digital selection (e.g., a remote control signal) of a channel to be presented on the television **102**. Additionally or alternatively, codes embedded with or otherwise broadcast with media content being presented via the STB **108** and/or the television **102** may be utilized by the program detector **212** to identify the presentation. As used herein, a code is an identifier that is transmitted with the media content for the purpose of identifying and/or tuning the corresponding media content. Codes may be carried in the audio, in the video, in metadata, in a vertical blanking interval, in a program guide, in content data, or in any other portion of the media content or the signal carrying the content. Additionally or alternatively, the program detector **212** can collect a signature representative of a portion of the media content. As used herein, a signature is a representation of some characteristic of the media content (e.g., a frequency spectrum of an audio signal). Collected signature(s) can be compared against a collection of signatures of known media content to identify the corresponding media content. The signature(s) can be collected by the program detector **212** and/or the program detector **212** can collect samples of the media content and export them to a remote site for generation of the signature(s). Irrespective of the manner in which the media content of the presentation is identified, the identification information is time stamped by the time stamper **208** and maybe stored in the memory **210**.

In the illustrated example of FIG. **2**, the output device **214** periodically and/or aperiodically exports the recorded data from the memory **214** to a data collection facility via a network (e.g., a local-area network, a wide-area network, a metropolitan-area network, the Internet, a digital subscriber line (DSL) network, a cable network, a power line network, a wireless communication network, a wireless mobile phone network, a Wi-Fi network, etc.). The data collection facility of the illustrated example utilizes the people tallies generated by the people counter **206** and/or the personal IDs generated by the person identifier **207** in conjunction with the content identifying data collected by the program detector **212** to generate exposure information and/or compliance information (e.g., indications of whether or not members of a panel have behaved in accordance with term(s) of membership). Alternatively, the data analysis could be performed locally and exported via a network or the like to a data collection facility for further processing. For example, the amount of people (as counted by the people counter **206**) in the exposure environment **100** at a time (as indicated by the time stamp appended to the people tally by the time stamper **208**) in which a sporting event (as identified by the program detector **212**) was presented by the television **102** can be used in a rating calculation for the sporting event. In some examples, additional information (e.g., demographic data associated with one or more people personally identified by the person identifier **207**, geographic data, etc.) is correlated with the exposure information at the data collection facility to expand the usefulness of the data collected by the example audience measurement device **104** of FIGS. **1** and/or **2**. The data collection facility of the illustrated example compiles data from many monitored exposure environments.

While an example manner of implementing the audience measurement device **104** of FIG. **1** has been illustrated in FIG. **2**, one or more of the elements, processes and/or devices illustrated in FIG. **2** may be combined, divided, re-arranged, omitted, eliminated and/or implemented in any other way. Further, the example audience detector **200**, the example content identifier **202**, the example image sensor **204**, the example people counter **206**, the person identifier **207**, the example time stamper **208** and/or, more generally, the example audience measurement device **104** of FIG. **2** may be implemented by hardware, software, firmware and/or any combination of hardware, software and/or firmware. Thus, for example, any of the example audience detector **200**, the example content identifier **202**, the example image sensor **204**, the example people counter **206**, the person identifier **207**, the example time stamper **208** and/or, more generally, the example audience measurement device **104** of FIG. **2** could be implemented by one or more circuit(s), programmable processor(s), application specific integrated circuit(s) (ASIC(s)), programmable logic device(s) (PLD(s)) and/or field programmable logic device(s) (FPLD(s)), field programmable gate array (FPGA), etc. When any of the apparatus or system claims of this patent are read to cover a purely software and/or firmware implementation, at least one of the example audience detector **200**, the example content identifier **202**, the example image sensor **204**, the example people counter **206**, the person identifier **207**, the example time stamper **208**, and/or the example audience measurement device **104** of FIG. **2** are hereby expressly defined to include a tangible computer readable medium such as a memory, DVD, CD, BluRay, etc. storing the software and/or firmware. Further still, the example audience measurement device **104** of FIG. **2** may include one or more elements, processes and/or devices in addition to, or instead of, those illustrated in FIG. **2**, and/or may include more than one of any or all of the illustrated elements, processes and devices.

FIG. **3** is a block diagram of an example implementation of the example image sensor **204** of FIG. **2**. The example image sensor **204** of FIG. **3** includes a resolution determiner **300**, a high-resolution (hi-res) trigger **302**, a frame cache **303**, an illumination controller **304**, an illuminator **306**, an image capturer **308**, a frame router **310**, and a pixel binner **312**. In the illustrated example, the image sensor **204** operates in a first mode for frames not designated for processing that requires high-resolution image data. The first mode is referred to herein as a majority mode. In some examples, a default mode of the image sensor **204** is the majority mode. The example image sensor **204** also operates in a second mode for frames designated for processing that requires high-resolution image data, such as frames designated for a facial recognition process to be executed by the example person identifier **207** of FIG. **2**. The second mode is referred to herein as a minority mode. The first mode is referred to as the majority mode and the second mode is referred to as a minority mode because the image sensor **204** is expected (but not necessarily) to be in the majority mode more often than the minority mode.

The resolution determiner **300** of the illustrated example determines in which mode the image sensor **204** is to operate for a given time period, an image frame and/or set of image

frames. In the illustrated example of FIG. **3**, the resolution determiner **300** causes the image sensor to operate in the majority mode unless the minority mode has been triggered. As described above, the minority mode is to be entered when a frame to be captured by the image capturer **308** is designated to undergo, for example, a facial recognition process. To determine when an upcoming frame or set of frames is to undergo such resolution-demanding process, the example hi-res trigger **302** references the frame cache **303**. The example frame cache **303** of FIG. **3** is a memory that stores frames recently captured by the image capturer **308**. The example frame cache **303** of FIG. **3** receives frames directly from the image capturer **308** and/or the pixel binner **312**, which is described in detail below. The hi-res trigger **302** retrieves one or more of the recently captured frames from the frame cache **303** and analyzes the frame(s) to determine whether a shape corresponding to a human head and/or face appears in the previous frame(s). In some examples, the hi-res trigger **302** also determines whether the detected head and/or face of the previous frames is oriented such that a facial recognition process can be successfully performed on the corresponding image data. In some examples, the hi-res trigger **302** also determines a period of time during which the detected head and/or face has been in a general location (e.g., if the corresponding person is not moving from a general position in the frame). Given the results of the one or more analyses performed by the hi-res trigger **302** on the previous frame(s), the example hi-res trigger **302** of FIG. **3** determines whether succeeding frame(s) are to undergo a facial recognition process to attempt to identify the detected head(s) and/or face(s). In some examples, the hi-res trigger **302** is a clock that drives the system into the minority mode at fixed intervals. For each frame that the hi-res trigger **302** designates as a frame that will be subject to a facial recognition process (either via analysis or based on a clock), the example hi-res trigger **302** of FIG. **3** places the image sensor **204** in the minority mode. Otherwise, the example image sensor **204** of FIG. **3** operates according to the majority mode.

When the example image sensor **204** of FIG. **3** is operating in the majority mode (e.g., the hi-res trigger **302** determines that facial recognition is not to be performed on the corresponding frame(s)), the resolution determiner **300** of the illustrated example conveys an instruction or indication to the illumination controller **304** that causes the illumination controller **304** to disable the illuminator for the corresponding frame(s). In the illustrated example, the illuminator **306** is an IR LED that emits IR light to illuminate the environment **100** with IR light detectable by the image capturer **308**, which includes IR capabilities. When in the majority mode, the example illumination controller **304** coordinates with the image capturer **308** to synchronize inaction of the illuminator **306** with the capture of the corresponding frame(s). Thus, the resolution determiner **300** causes the illuminator **306** to be inactive during capture of frames not designated for facial recognition processes (and/or only causes the illuminator **306** to be active during capture of frames that are designated for facial recognition processing. As described above, disabling the illuminator **306** reduces power consumption and extends the lifetime of the image sensor **204**. Further, disabling the illuminator **306** reduces the heat generated by the illuminator and, thus, reduces or even eliminates the need for heat sinking devices and/or techniques. Further, when the illuminator **306** is not blinking and/or glowing (e.g., when active), members of the audience **106** are less likely to become annoyed by the image sensor **204**.

When the example image sensor **204** of FIG. **3** is operating in the minority mode (e.g., the hi-res trigger **302** determines that facial recognition is to be performed on the corresponding frame(s)), the resolution determiner **300** conveys an instruction to the illumination controller **304** that causes the illumination controller **304** to enable the illuminator **306** during capture of the corresponding frame(s). The example illumination controller **304** coordinates with the image capturer **308** to synchronize operation of the illuminator **306** with the capture of the corresponding frame(s). Thus, the hi-res trigger **302** causes the illuminator **306** to be active for frames designated for facial recognition processes (e.g., by the person identifier **207**). Frames may be specifically designated or may be generally designated as "frames which occur at a certain time." Accordingly, when operating in the majority mode, the illuminator **306** does not illuminate the environment **100**. In contrast, when operating the minority mode, the illuminator **306** does illuminate the environment **100**.

The example image capturer **308** of FIG. **3** is a high-resolution camera that captures image data representative of the environment **100** of FIG. **1**. In the illustrated example, the image capturer **308** operates similarly in the majority mode as in the minority mode. That is, the example image capturer **308** of FIG. **3** captures high-resolution image data for each captured frame. Given the selective operation of the illuminator **306** described above, the high-resolution frames captured when operating in the majority mode are likely to be poorly lit. On the other hand, the high-resolution frames captured when operating in the minority mode are well lit. The example image capturer **308** of FIG. **3** conveys all or some (e.g., only the frames captured in the minority mode) of the captured frames to the resolution determiner **300** for storage in the frame cache **303**. As described above, the frames stored in the frame cache **303** are used by the example hi-res trigger **302** to identify heads and/or faces in the environment **100** that, due to their orientation, may be identifiable by the person identifier **207**.

Further, the example image capturer **308** of FIG. **3** conveys the captured high-resolution frames to the frame router **310**. The example frame router **310** of FIG. **3** also receives an instruction from the resolution determiner **300** indicative of the mode (e.g., majority or minority) in which the image sensor **204** is operating for the received frames. For frames designated by the resolution determiner **300** as frames captured in the minority mode, the illuminator **306** was active and, thus, the image data is of high-resolution and well lit. Accordingly, the image data captured in the minority mode is in condition for the facial recognition process. Therefore, in response to the indicator from the resolution determiner **300** that the received frame(s) were captured in the minority mode, the example frame router **310** of FIG. **3** routes the illuminated, high-resolution frames to the person identifier **207** for facial recognition process(es). In some examples, the frame router **310** also routes the high-resolution frames to the people counter **206** for counting process(es).

For frames designated by the resolution determiner **300** as frames captured in the majority mode, the illuminator **306** was inactive and, thus, the image data is of high-resolution but likely to be dark and of low contrast. For many of these frames, the contrast level is likely to be too low for proper analysis (e.g., people counting) to be accurately performed. Accordingly, the image data captured in the majority mode may not be in condition (e.g., may not have high enough contrast levels) for an accurate execution of the people counting process of the people counter **206**. Therefore, in

response to the instruction from the resolution determiner **300** that the received frame(s) were captured in the majority mode, the example frame router **310** of FIG. **3** routes the low-contrast frames to the pixel binner **312**.

The example pixel binner **312** of FIG. **3** receives high-resolution frames of image data from the frame router **310** that are likely to be of low contrast levels due to the illuminator **306** being inactive during capture of the frames. The example pixel binner **312** of FIG. **3** performs a binning procedure on the received frames. Binning is a process in which a neighborhood of pixels (e.g., a first pixel and a number of adjacent pixels) is summed together. Example binning procedures sum pixels according to a 2×2 neighborhood of pixels, a 3×3 neighborhood of pixels, a 4×4 neighborhood of pixels, or another arrangement of pixels. In effect, the binning procedure performed by the pixel binner **312** may assign more light per orthogonal block of pixels at the cost of lower resolution. The additional light per block pixel increases the contrast levels of the image data. After the pixel procedure performed by the example pixel binner **312** of FIG. **3**, the frames are of lower resolution (e.g., in comparison with the original format of the frame as captured by the high-resolution camera of the image capturer **308**) but may also have higher contrast levels. Thus, the example pixel binner **312** enhances the dark, low-contrast images generated in the majority mode for processing that does not require high-resolution image data (e.g., the facial recognition process of the person identifier **207**), such as the people counting process of the people counter **206**. The example pixel binner **312** conveys the resulting frames of image data to the people counter **206**. The higher contrast levels provided by the example pixel binner **312** may enable the people counter **206** to distinguish between shapes in the frames of image data. The example pixel binner **312** of FIG. **3** also conveys the frames of image data to the resolution determiner **300** for storage in the frame cache **303**.

While an example manner of implementing the image sensor **204** of FIG. **2** has been illustrated in FIG. **3**, one or more of the elements, processes and/or devices illustrated in FIG. **3** may be combined, divided, re-arranged, omitted, eliminated and/or implemented in any other way. Further, the example resolution determiner **300**, the example hi-res trigger **302**, the example illumination controller **304**, the example frame router **310**, the example pixel binner **310**, and/or, more generally, the example image sensor **204** of FIG. **3** may be implemented by hardware, software, firmware and/or any combination of hardware, software and/or firmware. Thus, for example, any of the example resolution determiner **300**, the example hi-res trigger **302**, the example illumination controller **304**, the example frame router **310**, the example pixel binner **310**, and/or, more generally, the example image sensor **204** of FIG. **3** could be implemented by one or more circuit(s), programmable processor(s), application specific integrated circuit(s) (ASIC(s)), programmable logic device(s) (PLD(s)) and/or field programmable logic device(s) (FPLD(s)), field programmable gate array (FPGA), etc. When any of the apparatus or system claims of this patent are read to cover a purely software and/or firmware implementation, at least one of, the example resolution determiner **300**, the example hi-res trigger **302**, the example illumination controller **304**, the example frame router **310**, the example pixel binner **310**, and/or the example image sensor **204** of FIG. **3** are hereby expressly defined to include a tangible computer readable medium such as a memory, DVD, CD, Bluray, etc. storing the software and/or firmware. Further still, the example image sensor **204** of FIG. **3** may include one or more elements, processes and/or devices in addition to, or instead of, those illustrated in FIG. **3**, and/or may include more than one of any or all of the illustrated elements, processes and devices.

FIG. **4** is a flowchart representative of example machine readable instructions for implementing the example image sensor **204** of FIGS. **1**, **2** and/or **3**. In this example, the machine readable instructions comprise a program for execution by a processor such as the processor **512** shown in the example processing system **500** discussed below in connection with FIG. **5**. The program may be embodied in software stored on a tangible computer readable medium such as a CD-ROM, a floppy disk, a hard drive, a digital versatile disk (DVD), a BluRay disk, or a memory associated with the processor **512**, but the entire program and/or parts thereof could alternatively be executed by a device other than the processor **512** and/or embodied in firmware or dedicated hardware. Further, although the example program is described with reference to the flowchart illustrated in FIG. **4**, many other methods of implementing the example image sensor **204** may alternatively be used. For example, the order of execution of the blocks may be changed, and/or some of the blocks described may be changed, eliminated, or combined.

As mentioned above, the example processes of FIG. **4** may be implemented using coded instructions (e.g., computer readable instructions) stored on a tangible computer readable medium such as a hard disk drive, a flash memory, a read-only memory (ROM), a compact disk (CD), a digital versatile disk (DVD), a cache, a random-access memory (RAM) and/or any other storage media in which information is stored for any duration (e.g., for extended time periods, permanently, brief instances, for temporarily buffering, and/or for caching of the information). As used herein, the term tangible computer readable medium is expressly defined to include any type of computer readable storage and to exclude propagating signals. Additionally or alternatively, the example processes of FIG. **4** may be implemented using coded instructions (e.g., computer readable instructions) stored on a non-transitory computer readable medium such as a hard disk drive, a flash memory, a read-only memory, a compact disk, a digital versatile disk, a cache, a random-access memory and/or any other storage media in which information is stored for any duration (e.g., for extended time periods, permanently, brief instances, for temporarily buffering, and/or for caching of the information). As used herein, the term non-transitory computer readable medium is expressly defined to include any type of computer readable medium and to exclude propagating signals. As used herein, when the phrase "at least" is used as the transition term in a preamble of a claim, it is open-ended in the same manner as the term "comprising" is open ended. Thus, a claim using "at least" as the transition term in its preamble may include elements in addition to those expressly recited in the claim.

FIG. **4** begins with an initiation of the example image sensor **204** of FIG. **3** (block **400**). In the illustrated example, the initiation of the image sensor **204** corresponds to an initiation of a monitoring session of the example exposure environment **100** of FIG. **1**. The example resolution determiner **300** determines whether the next frame(s) to be captured are to be subjected to processing that demands high-resolution image data (block **402**). In the illustrated example, the amount of frames for which the determination is to be made varies depending on, for example, an adjustable setting and/or a finding of the hi-res trigger **302** regarding previous frame(s). As described above, the determination of the resolution determiner **300** indicates which mode (e.g., majority (e.g., low resolution) mode or minority

(e.g., high resolution) mode) in which the image sensor **204** is to operate for the next frame(s). In the illustrated example, the resolution determiner **300** bases the determination on whether the next frame(s) to be captured will undergo a facial recognition process implemented by the example person identifier **207** of FIG. **2**. For example, the hi-res trigger **302** of FIG. **3** determines whether a head or face is included in previous frame(s) (e.g., frames(s) obtained from the frame cache **303**) capable of providing an identification of a corresponding person.

In the example of FIG. **4**, when the resolution determiner **300** determines that the next frame(s) to be captured are not going to be subjected to the facial recognition process (block **402**), the example resolution determiner **300** provides an instruction to the illumination controller **304** to disable the illuminator **306** during capture of the next frame(s) (block **404**). In such instances, the illumination controller **304** cooperates with the image capturer **308** to capture the next frame(s) with the illuminator **306** deactivated (block **406**). While the resulting frames are captured with the high-resolution camera of the image capturer **308**, the frames are not lit by the illuminator **306** and, thus, likely to be dark and to have low contrast levels. To increase the contrast levels of the frame(s) captured without use of the illuminator **306**, the frame router **310** routes the frame(s) to the pixel binner **312** (block **408**). The example frame router **310** of FIG. **3** is aware of which frames were captured without use of the illuminator **306** via the instruction provided thereto by the resolution determiner **300** regarding in which mode (e.g., majority mode or minority mode) the image sensor **204** is operating.

The example pixel binner **312** of FIG. **3** performs a binning procedure (e.g., a 2×2 binning) on the received frame(s) (block **410**). As described above, the binning procedure increases the contrast levels of the image data at the expense of the resolution of the image data. The lower resolution is acceptable for the frame(s) captured while in the majority mode because the people counting process(es) for which the majority mode frames are designated does not require high-resolution data, but operates more accurately with higher contrast levels. Having performed the binning procedure on the majority mode frames, the example pixel binner **312** conveys the binned frame(s) to the people counter **206** and the frame cache **303** (block **412**). Control returns to block **402**.

Referring back to block **402**, when the resolution determiner **300** determines that the next frame(s) to be captured are going to be subjected to the facial recognition process, the example resolution determiner **300** provides an instruction to the illumination controller **304** to enable the illuminator **306** during capture of the next frame(s) (block **414**). In such instances, the illumination controller **304** cooperates with the image capturer **308** to capture the next frame(s) with the illuminator **306** activated (e.g., emitting light, such as IR light into the exposure environment **100** of FIG. **1**) (block **416**). The resulting frames are captured with the high-resolution camera of the image capturer **308** and the frames are lit by the illuminator **306**. Thus, the minority mode frames are ready for processing by the person identifier **207**. Accordingly, the route framer **310** conveys the minority mode frames to the person identifier **207** and the frame cache **303** (block **418**). Control returns to block **402**.

While the example image sensor of FIGS. **1**, **2** and/or **3** is described in the context of an audience measurement device **104** and the generation of exposure data for media content, the example methods, apparatus, and articles of manufacture disclosed herein can be applied to additional or alternative contexts, systems, measurements, applications, programs, etc. That is, the example methods, apparatus, and articles of manufacture disclosed herein can be used in any application to selectively operate an illumination source of an image sensor based on a resolution demand for the resulting frames of image data.

FIG. **5** is a block diagram of an example processor system **510** capable of executing the instructions of FIG. **4** to implement the example image sensor **204** of FIGS. **2** and/or **3**. The processor system **510** can be, for example, a server, a personal computer, a mobile phone, a personal digital assistant (PDA), an Internet appliance, a DVD player, a CD player, a digital video recorder, a BluRay player, a gaming console, a personal video recorder, a set-top box, an audience measurement device, or any other type of computing device.

The example processor system **510** of FIG. **5** includes a processor **512** that is coupled to an interconnection bus **514**. The processor **512** may be any suitable processor, processing unit, or microprocessor (e.g., one or more Intel® microprocessors from the Pentium® family, the Itanium® family or the XScale® family and/or other processors from other families). The system **510** may be a multi-processor system and, thus, may include one or more additional processors that are identical or similar to the processor **512** and that are communicatively coupled to the interconnection bus **514**.

The processor **512** of FIG. **5** is coupled to a chipset **518**, which includes a memory controller **520** and an input/output (I/O) controller **522**. A chipset provides I/O and memory management functions as well as a plurality of general purpose and/or special purpose registers, timers, etc. that are accessible or used by one or more processors coupled to the chipset **518**. The memory controller **520** performs functions that enable the processor **512** to access a system memory **524**, a mass storage memory **525**, and/or a digital versatile disk (DVD) **540**.

In general, the system memory **524** may include any desired type of volatile and/or non-volatile memory such as, for example, static random access memory (SRAM), dynamic random access memory (DRAM), flash memory, read-only memory (ROM), double data rate memory (DDR), etc. The mass storage memory **525** may include any desired type of mass storage device including hard disk drives, optical drives, tape storage devices, etc. The machine readable instructions of FIGS. **5A-5C** may be stored in the system memory **524**, the mass storage memory **525**, and/or the DVD **540**.

The I/O controller **522** performs functions that enable the processor **512** to communicate with peripheral input/output (I/O) devices **526** and **528** and a network interface **530** via an I/O bus **532**. The I/O devices **526** and **528** may be any desired type of I/O device such as, for example, a keyboard, a video display or monitor, a mouse, etc. The network interface **530** may be, for example, an Ethernet device, an asynchronous transfer mode (ATM) device, an 802.11 device, a digital subscriber line (DSL) modem, a cable modem, a cellular modem, etc. that enables the processor system **510** to communicate with another processor system. The example network interface **530** of FIG. **5** is also communicatively coupled to a network **534**, such as an intranet, a Local Area Network, a Wide Area Network, the Internet, etc.

While the memory controller **520** and the I/O controller **522** are depicted in FIG. **5** as separate functional blocks within the chipset **518**, the functions performed by these

blocks may be integrated within a single semiconductor circuit or may be implemented using two or more separate integrated circuits.

Although certain example apparatus, methods, and articles of manufacture have been disclosed herein, the scope of coverage of this patent is not limited thereto. On the contrary, this patent covers all apparatus, methods, and articles of manufacture fairly falling within the scope of the claims of this patent.

What is claimed is:

1. An audience measurement system to obtain exposure data for a media exposure environment, the audience measurement system comprising:
- memory;
- machine readable instructions; and
- processor circuitry to execute the machine readable instructions to:
  - generate an audio signature of media content presented by a television within the media exposure environment;
  - obtain content identifying data corresponding to the presented media content, the content identifying data based on the audio signature of the media content presented by the television within the media exposure environment;
  - analyze a sequence of images of the media exposure environment to detect a head appearing in one or more of the images, the sequence of images obtained by a camera while the media content corresponding to the content identifying data is presented by the television;
  - determine an orientation of the head with respect to the camera; and
  - determine audience identification information based on a match of the head to a known person associated with the media exposure environment; and
- network interface circuitry to output a signal indicative of the content identifying data and the audience identification information to a data collection facility.

2. The audience measurement system of claim **1**, wherein the audio signature is a representation of a frequency spectrum of an audio signal of the media content.

3. The audience measurement system of claim **1**, wherein the signal is a first signal; the processor circuitry is to determine a people tally indicative of a count of people appearing within the one or more of the images based at least in part on the detected head; and the network interface circuitry is to output a second signal indicative of the people tally.

4. The audience measurement system of claim **1**, wherein the processor circuitry is to:
- reduce a resolution of a first image of the one or more of the images of the media exposure environment to obtain a reduced-resolution image, and
- determine the orientation of the head with respect to the camera based on the reduced-resolution image.

5. The audience measurement system of claim **4**, wherein the processor circuitry is to determine the audience identification information by: (i) generating a facial signature from a region of a second image of the one or more of the images corresponding to a location of the head in the reduced-resolution image, and (ii) comparing the generated facial signature to a database of facial signatures.

6. The audience measurement system of claim **4**, wherein the processor circuitry is to:
- identify a region corresponding to the head within the first image from which the reduced-resolution image was obtained;
- analyze a portion corresponding to the identified region of a second image; and
- based on the analysis of the region of the second image, determine that the head matches the known person.

7. The audience measurement system of claim **6**, wherein the processor circuitry is to identify a region corresponding to the head within the first image from which the reduced-resolution image was obtained by determining that a face is potentially identifiable based on the orientation of the head.

8. The audience measurement system of claim **6**, wherein the audience identification information includes an identifier associated with the known person and the processor circuitry is to, responsive to a determination that the head matches the known person, cause the identifier associated with the person to be stored in the memory.

9. An audience measurement system, comprising:
- an audience measurement device at a media exposure environment, the audience measurement device to:
  - generate an audio signature of media content presented by a television within the media exposure environment;
  - obtain content identifying data corresponding to the presented media content, the content identifying data based on the audio signature;
  - while the media content corresponding to the content identifying data is presented by the television, collect a first image of the media exposure environment with a camera;
  - attempt to detect a head in the first image;
  - determine an orientation of the head with respect to the camera;
  - determine audience identification information based on an indication that the head matches a known person associated with the audience measurement system; and
- a data collection facility to obtain the content identifying data and the audience identification information from the audience measurement device.

10. The audience measurement system of claim **9**, wherein the audio signature is a representation of a frequency spectrum of an audio signal of the media content.

11. The audience measurement system of claim **9**, wherein the audience measurement device is to:
- reduce a resolution of the first image of the media exposure environment to obtain a reduced-resolution image, and
- determine the orientation of the head with respect to the camera based on the reduced-resolution image.

12. The audience measurement system of claim **11**, wherein the audience measurement device is to:
- identify a region corresponding to the head within the first image from which the reduced-resolution image was obtained; and
- analyze a portion of a second image corresponding to the identified region to determine whether the head matches the known person.

13. The audience measurement system of claim **12**, wherein the audience identification information includes an identifier associated with the known person and the audience measurement device is to, responsive to the indication that the head matches the known person, cause recordation of the identifier associated with the known person in a memory.

14. The audience measurement system of claim **11**, wherein the audience measurement device is to determine the audience identification information by: (i) generating a facial signature from a region of a second image corresponding to a location of the head in the reduced-resolution image, and (ii) comparing the generated facial signature to a database of facial signatures.

15. The audience measurement system of claim **9**, wherein the audience measurement device is to:

determine a people tally indicative of a count of people appearing in the first image based at least in part on the detected head; and

provide the people tally to the data collection facility.

16. A method for obtaining exposure data for a media exposure environment, the method comprising:

generating, using a program detector, an audio signature of media content presented by a television within the media exposure environment;

obtaining, based on the audio signature, content identifying data corresponding to the presented media content;

while the media content corresponding to the content identifying data is presented by the television, capturing first and second images of the media exposure environment with a camera;

analyzing, using an audience detector, the first image to attempt to detect a head in the first image;

determining, using the audience detector, an orientation of the head with respect to the camera;

determining, using the audience detector, audience identification information based on an indication that the head matches a known person associated with the audience detector; and

providing the content identifying data and the audience identification information to a data collection facility.

17. The method of claim **16**, wherein the audio signature is a representation of a frequency spectrum of an audio signal of the media content.

18. The method of claim **16**, further comprising:

reducing a resolution of the first image of the media exposure environment to obtain a reduced-resolution image, and

wherein the determining of the orientation of the head includes determining the orientation of the head with respect to the camera using the reduced-resolution image.

19. The method of claim **18**, further comprising:

identifying a region corresponding to the head within the first image from which the reduced-resolution image was obtained;

analyzing a portion of the second image corresponding to the identified region; and

based on the analyzing of the region of the second image, determining whether the head matches the known person.

20. The method of claim **19**, wherein the audience identification information includes an identifier associated with the known person, and further including recording the identifier associated with the known person in a memory.

21. The method of claim **16**, further comprising:

determining a people tally indicative of a count of people appearing within the first image based at least in part on the detected head and one or more additional heads; and

providing the people tally to the data collection facility.

22. An audience measurement system to obtain exposure data for a media exposure environment, the audience measurement system comprising:

memory;

processor circuitry; and

program detector software to cause the processor circuitry to:

generate an audio signature of media content presented by a television within the media exposure environment, and

obtain, based on the audio signature, content identifying data corresponding to the presented media content;

audience detector software to cause the processor circuitry to:

while the media content corresponding to the content identifying data is presented by the television, obtain first and second images of the media exposure environment captured with a camera;

analyze one or more of the first or second images to: (i) detect one or more heads in the media exposure environment; (ii) determine corresponding orientations with respect to the camera for respective ones of the one or more heads, and (iii) determine audience identification information indicative that one or more of the heads matches respective ones of one or more persons known to be associated with the media exposure environment; and

network interface circuitry to output the content identifying data and the audience identification information to a data collection facility.

23. The audience measurement system of claim **22**, wherein the audience detector software is to cause the processor circuitry to analyze the one or more of the first or second images to detect the one or more heads in the media exposure environment by:

analyzing the first image to detect the one or more heads in the media exposure environment, and

analyzing the second image to determine the audience identification information.

24. The audience measurement system of claim **23**, wherein the audience detector software is to cause activation of an illumination source while the camera captures the second image.

25. The audience measurement system of claim **24**, wherein the audience detector software is to cause the camera to capture the first image without use of the illumination source.

26. The audience measurement system of claim **23**, wherein:

the audience detector software is to cause the processor circuitry to reduce a resolution of the first image using pixel binning to obtain a reduced-resolution image, and

to detect the one or more heads in the media exposure environment based on the reduced-resolution image.

27. The audience measurement system of claim **22**, wherein the audio signature is a representation of a frequency spectrum of an audio signal of the media content.

28. The audience measurement system of claim **22**, wherein the audience detector software is to cause the processor circuitry to determine the audience identification information by: (i) generating a facial signature from a region of the second image corresponding to a location of one of the one or more heads, and (ii) comparing the generated facial signature to a database of facial signatures.

29. The audience measurement system of claim **22**, wherein the audience detector software is to determine a people tally indicative of a count of people appearing within the first image based at least in part on the detected one or more heads, and

the network interface circuitry is to provide the people tally to the data collection facility.

* * * * *

EXHIBIT B

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF DELAWARE

THE NIELSEN COMPANY (US), LLC,	)	
	)	
Plaintiff,	)	
	)	C.A. No. ___________________
v.	)	
	)	**JURY TRIAL DEMANDED**
TVISION INSIGHTS, INC.,	)	
	)	
Defendant.	)	

DECLARATION OF PIERRE MOULIN

I, Pierre Moulin, declare as follows:

1. I have been retained by The Nielsen Company (US), LLC ("Nielsen") to provide an independent opinion regarding U.S. Patent No. 11,470,243 ("the '243 Patent").

2. I am not an employee of Nielsen or any affiliate or subsidiary of Nielsen. Nor do I have a financial interest in Nielsen or the outcome of this case. I am being compensated for my work at my standard consulting rate. My compensation is not dependent on the outcome of this case or on the content of my opinions.

3. The '243 Patent relates to audience measurement, including the identification of audience members and the determination of the orientation of their heads. In my opinion, a person of ordinary skill in the art in the field of the '243 Patent would have a working knowledge of basic computer technology and video processing technology. The person would gain this knowledge through either (i) an undergraduate Bachelor of Science degree in Computer Science or Electrical Engineering or a comparable field, and at least three years of work experience in relevant fields; or (ii) a Master's degree in Computer Science or Electrical Engineering and at least one year of work experience in relevant fields. I base these statements on my experience, including my knowledge of colleagues and others at the time of the invention of the '243 Patent.

4. Unless otherwise noted, the statements made in this Declaration are based on my personal knowledge, and if called to testify about this declaration, I could and would do so competently and truthfully.

5. A detailed record of my professional qualifications is attached to this Declaration as Exhibit A.

6. I have reviewed the '243 Patent and other materials that are listed in Exhibit B to this Declaration. My opinions are, in part, a result of my review of those materials. My opinions are also a result of my years of experience in the field. Moreover, my experience has informed my choices of materials to review.

7. I am not a legal expert and offer no opinions on the law.

8. I am a Professor at the University of Illinois ("UIUC") in the Department of Electrical and Computer Engineering, a Research Professor in the Coordinated Science Laboratory, a faculty member in the Beckman Institute for Advanced Science and Technology and its Image Formation and Processing Group, and an affiliate professor in the University's Department of Statistics. I am also a member of the Information Trust Institute and the founding director of the Center for Information Forensics, a multidisciplinary research center.

9. I am an experienced researcher and educator in the field of computer engineering, with expertise in signal processing, audio and video processing, information theory, machine learning, and large data systems.

10. I have experience and expertise in numerous areas including information theory, audio and video processing, statistical signal processing and modeling, computer vision, decision theory, information hiding and authentication, and the application of multi-resolution signal analysis, optimization theory, and fast algorithms to these areas. As part of my professional

research, I have studied automated content filtering, commonly referred to as digital "fingerprinting" or "signature generation" as it applies to both video and audio electronic content.

11. I have more than 35 years of experience in the field of electrical and computer engineering since receiving my bachelor's degree in 1984. I have 25 years of experience as a professor in the Electrical and Computer Engineering Department of UIUC.

12. During this time, I have led, overseen, and contributed to numerous research projects involving technical concepts that are closely related to the technology at issue in the '243 Patent. Full details of those projects are set forth in Exhibit A.

13. I have taught numerous courses on technology relating to audio and video processing, digital signal processing, information theory, and detection and estimation theory. I have developed courses in image and video processing, machine learning, and large data computing.

14. In my career field, I have received numerous accolades for my contributions to electrical and computer engineering, including being elected as Fellow of the IEEE in 2003. In 2018, I was honored to receive the Ronald W. Pratt Faculty Outstanding Teaching Award. I served on the IEEE Signal Processing Society Board of Governors from 2005-2007 and the IEEE Information Theory Society Board of Governors from 2016-2018. I also served as IEEE Signal Processing Society Distinguished Lecturer from 2012-2013.

15. I received my D.Sc. in Electrical Engineering (Washington University, St. Louis, 1990); M.Sc. in Electrical Engineering (Washington University, St. Louis, 1986); and Ingénieur Civil Electricien (Faculté Polytechnique de Mons, Belgium, 1984). The D.Sc. is equivalent to a Ph.D. and is recognized as such within academic and research communities.

16. After receiving my D.Sc. degree, I worked as a Research Scientist at Bell Communications Research, Morristown, NJ from 1990-1995.

17. My work as an academic began in 1996, when I joined the University of Illinois as Assistant Professor. In 1999, I was promoted to Associate Professor, and, in 2003, I was promoted to the position of Professor of Electrical and Computer Engineering and Statistics.

18. During sabbaticals from University of Illinois, I have held visiting positions at the National Taiwan University (2019), the Chinese University of Hong Kong (2010 and 2014), MIT (2005), and Microsoft Research (2001).

19. I have received funding from numerous agencies, foundations, and companies for my research on electrical and computer engineering, including image and signal processing. The sources of funding for this research include the National Science Foundation (NSF), the Defense Advanced Research Projects Agency (DARPA), Microsoft, HP Labs, Xerox, and the North Atlantic Treaty Organization (NATO).

20. I have served in a number of leadership roles for the IEEE, including as a Senior Editorial Board member for the IEEE Journal of Selected Topics in Signal Processing (2019-present), as an Editorial Board member for the *Proceedings of the IEEE* (2007-2012), as Co-founder and Editor-in-chief for *IEEE Transactions on Information Forensics and Security* (2005-2008), and as Area Editor for *IEEE Transactions on Image Processing* (2002-2006).

21. I have been an invited lecturer at academic institutions and organizations including MIT, Stanford, Berkeley, Carnegie-Mellon, Harvard, Bell Labs, IBM Research, and Microsoft Research, among others.

22. I have graduated twelve Ph.D. students and educated 9 post-doctoral students. Three of my Ph. D. students became associate professors upon graduation, and one became a

professor upon graduation. My other Ph.D. students have joined companies such as Intel Research, Qualcomm, and Google. Seven of my post-doctoral students began careers in academia in professorial positions.

23. I have researched and written about video and audio processing, video and audio hashing, video and audio identification, large data systems, and other topics related to video and audio processing.

24. During the course of my career, I have authored or co-authored more than 65 scientific articles, one book, and six book chapters. I have presented and/or published 194 conference papers. I have also been awarded two U.S. patents. These, and other aspects of my qualifications, are summarized in my curriculum vitae accompanying this declaration as Exhibit A.

25. The '243 Patent concerns, among other things, particular approaches for measuring media audiences. (*See*, e.g., '243 Patent, Complaint Ex. A, Claims 4-6, 8, 11-14, 18-20.) Measurement is accomplished through, among other things, specific methods for determining information about audience members by comparing data about the members' heads or faces to known reference information. (*See id.*)

26. In certain prior art audience measurement systems, a series of images of the media exposure environment (*e.g.*, a living room with a television) is collected over a period of time. From that series of images, those systems determine information about audience members present. ('243 Patent, Complaint Ex. A, 1:56-2:3.) Because the collected images are taken over a period of time, information about audience members can be tracked over time and correlated to particular frames or segments of the media content being displayed on the television.

27. To determine the identities of audience members, certain prior art audience measurement systems have used facial recognition. ('243 Patent, Complaint Ex. A, 2:42-56.) To accurately perform facial recognition analysis on a captured image of a media exposure environment, the image must have both sufficiently high resolution (*i.e.*, a sufficiently large density of pixels) and sufficiently high contrast (*i.e.*, sufficiently different brightness levels between audience members and their surroundings). ('243 Patent, Complaint Ex. A, 2:42-3:3.) To obtain such an image, the media exposure environment must have sufficient lighting. ('243 Patent, Complaint Ex. A, 2:57-3:3.) Because television rooms typically have poor ambient lighting, some audience measurement systems employ their own illumination source every time an image of sufficiently high resolution and sufficiently high contrast is to be collected. ('243 Patent, Complaint Ex. A, 2:57-3:3.)

28. There are several disadvantages of having to ensure that the media exposure environment is well-illuminated. The most apparent disadvantage of a system that requires substantial illumination all the time is that if the environment is poorly-lit, it may simply be impossible to collect satisfactory images. Or, in such a poorly-lit environment, an illumination source may be required. But even if such an illumination source is available, there are numerous disadvantages. First, use of an illumination source is a significant power drain. Second, frequent use of an illumination source necessitates its frequent replacement. Third, an illumination source creates excess heat that must be controlled with heat sinking devices. And fourth, an illumination source tends to annoy people, which in turn creates problems maintaining panelist participation. ('243 Patent, Complaint Ex. A, 3:4-30.)

29. The '243 Patent discloses that for operations other than facial recognition, it is not necessary to have a high-resolution image. Instead, a reduced-resolution image may be

used for actions such as determining the location and head-orientation of audience members. However, the image used for those actions must have sufficiently high contrast. ('243 Patent, Complaint Ex. A, 3:31-4:14.) To that end, the '243 Patent discloses that a high-contrast, low-resolution image (sufficient for actions such as determining the location and head-orientation of audience members) can be taken without having to ensure a well-illuminated media exposure environment. ('243 Patent, Complaint Ex. A, 3:31-4:14.) In particular, the '243 Patent discloses that in exchange for loss of resolution, techniques such as "pixel binning" can be used to increase the contrast of images that are not well-illuminated. ('243 Patent, Complaint Ex. A, 3:31-62.)

30. The '243 Patent discloses that because the vast majority of the captured images of the media exposure environment are not used for facial recognition, it is not necessary to have a well-illuminated environment for capturing the vast majority of images (or all images, if accurate facial recognition is not required). For this reason, the '243 Patent refers to capturing images in a potentially poorly-illuminated environment as "majority capture mode," and it refers to capturing images in a necessarily well-illuminated environment as "minority capture mode." ('243 Patent, Complaint Ex. A, 3:63-4:14.)

31. Using low-resolution images for activities other than facial recognition carries the benefit of not needing a well-lit environment. This, in turn, also carries the benefit of minimizing (or eliminating, if accurate facial recognition is not required) the use of any illumination source, which in turn has the benefit of saving power, reducing heat, reducing illumination source replacement costs, and reducing panelist annoyance. ('243 Patent, Complaint Ex. A, 3:4-30.) Furthermore, the use of low-resolution images for activities other than facial recognition decreases the system's burdens of computation and storage. These burdens are decreased with

lower resolution because lower-resolution images have fewer pixels from which data is analyzed and stored. For example, if a full-resolution image contains 400,000 pixels, a reduced-resolution image might instead use groups of four pixels to form pixel blocks, effectively resulting in a 100,000-pixel reduced-resolution image. And finally, an additional benefit of using low-resolution resolution images for activities other than facial recognition is that low-resolution images may be formed by applying contrast-increasing processing (such as pixel binning) to higher-resolution images, resulting in higher-contrast images. This increased contrast allows improved determination of head detection, location, and orientation because it provides for easier distinction between the head and its surroundings.

32. Claim 4 of the '243 Patent (which incorporates the limitations of claim 1) recites analyzing a sequence of images of a media exposure environment obtained by a camera to detect the head of an audience member. The claim further recites reducing the resolution of a first image of the sequence of images and determining the orientation of the audience member's head with respect to the camera, based on the reduced-resolution image. Determining the orientation of an audience member's head based on a reduced-resolution image was not well-understood, routine, or conventional as of December 15, 2011 (the filing date of the parent application from which the '243 Patent issued).

33. An advantage over the prior art of determining the orientation of an audience member's head based on a reduced-resolution image, as recited in claim 4, is that a reduced-resolution image can be obtained from an image collected in a poorly-illuminated environment. This is so because, as explained above, a poorly-illuminated image can undergo processing that reduces its resolution in exchange for an increase in contrast level. A further advantage over the prior art of determining the orientation of an audience

member's head based on a reduced-resolution image is that processing such images requires fewer computational and storage resources than determining head orientation based on a higher-resolution image (such as a high-resolution image that would be suitable for facial recognition analysis). This is so, as explained above, because a reduced-resolution image contains fewer pixels from which data must be processed and stored. And finally, another advantage over the prior art of determining the orientation of an audience member's head based on a reduced-resolution image is that low-resolution images may be formed by applying contrast-increasing processing (such as pixel binning) to higher-resolution images, resulting in higher-contrast images. This increased contrast allows improved determination of head orientation because it provides for easier distinction between the head and its surroundings.

34. Claim 5 of the '243 Patent incorporates the limitations of claim 4, and thus, its approach has the same advantages over the prior art as claim 4.

35. In addition, claim 5 of the '243 Patent recites generating a facial signature from a region of a second (higher-resolution) image corresponding to the location of the head in the reduced-resolution image, and comparing the generated facial signature to a database of facial signatures. In other words, the claim recites finding the location of the head in the first image, and using that location in the second image (which is higher-resolution than the first image) for the generation of a facial signature. That facial signature is compared to reference facial signatures for identification of the audience member. The approach of this claim was not well-understood, routine, or conventional in the prior art (*i.e.*, as of December 15, 2011). As compared to the prior art, the approach of claim 5 is an improvement because it employs reduced-resolution images where possible (*i.e.*, when locating the head). In turn, as explained

above, the advantages of using reduced-resolution images are (1) the ability to accomplish most operations in a poorly-illuminated environment; (2) reduction in computational and storage burdens (due to fewer pixels being processed and stored); and (3) increased contrast resulting from the ability to form low-resolution images by applying contrast-increasing processing to higher-resolution images (resulting in improved head location and orientation determination).

36. Claim 6 of the '243 Patent incorporates the limitations of claim 4, and thus, its approach has the same advantages over the prior art as claim 4.

37. In addition, claim 6 of the '243 Patent recites identifying a region corresponding to the head within the first image from which the reduced-resolution image was obtained, analyzing a portion corresponding to the identified region in a second (higher-resolution) image, and determining that the head matches a known person. In other words, claim 6 recites finding the location of the audience member's head in the reduced-resolution image, and using that location in a higher-resolution image for an analysis that allows determination of the identity of the audience member. The approach of this claim was not well-understood, routine, or conventional in the prior art. As compared to the prior art, the approach of claim 6 is an improvement because it employs reduced-resolution images where possible (*i.e.*, when locating the head). In turn, as explained above, the advantages of using reduced-resolution images are (1) the ability to accomplish most operations in a poorly-illuminated environment; (2) reduction in computational and storage burdens (due to fewer pixels being processed and stored); and (3) the opportunity to form low-resolution images through the use of contrast-increasing processing.

38. Claim 8 of the '243 Patent incorporates the limitations of claim 6, and thus, its approach has the same advantages over the prior art as claim 6.

39. Claim 11 of the '243 Patent (which incorporates the limitations of claim 9) recites a system that analyzes a first image of a media exposure environment collected by a camera to detect the head of an audience member. The claim further recites that the system reduces the resolution of the first image and determines the orientation of the audience member's head with respect to the camera, based on the reduced-resolution image. Determining the orientation of an audience member's head based on a reduced-resolution image was not well-understood, routine, or conventional as of December 15, 2011 (the filing date of the parent application from which the '243 Patent issued).

40. An advantage over the prior art of determining the orientation of an audience member's head based on a reduced-resolution image, as recited in claim 11, is that a reduced-resolution image can be obtained from an image collected in a poorly-illuminated environment. This is so because, as explained above, a poorly-illuminated image can undergo processing that reduces its spatial resolution in exchange for an increase in contrast level. A further advantage over the prior art of determining the orientation of an audience member's head based on a reduced-resolution image is that processing such images requires fewer computational and storage resources than determining head orientation based on a higher-resolution image (such as a high-resolution image that would be suitable for facial recognition analysis). This is so, as explained above, because a reduced-resolution image contains fewer pixels from which data must be processed and stored. And finally, another advantage over the prior art of determining the orientation of an audience member's head based on a reduced-resolution image is that low-resolution images may be formed by applying contrast-increasing processing (such as pixel binning) to higher-resolution images, resulting in higher-contrast images. This increased contrast

allows improved determination of head orientation because it provides for easier distinction between the head and its surroundings.

41. Claim 12 of the '243 Patent incorporates the limitations of claim 11, and thus, its approach has the same advantages over the prior art as claim 11.

42. In addition, claim 12 of the '243 Patent recites identifying a region corresponding to the head within the first image from which the reduced-resolution image was obtained, analyzing a portion corresponding to the identified region in a second (higher-resolution) image, and determining whether the head matches a known person. In other words, claim 12 recites finding the location of the audience member's head in the reduced-resolution image, and using that location in a higher-resolution image for an analysis that allows determination of the identity of the audience member. The approach of this claim was not well-understood, routine, or conventional in the prior art. As compared to the prior art, the approach of claim 12 is an improvement because it employs reduced-resolution images where possible (*i.e.*, when locating the head). In turn, as explained above, the advantages of using reduced-resolution images are (1) the ability to accomplish most operations in a poorly-illuminated environment; (2) reduction in computational and storage burdens (due to fewer pixels being processed and stored); and (3) the opportunity to form low-resolution images through the use of contrast-increasing processing.

43. Claim 13 of the '243 Patent incorporates the limitations of claim 12, and thus, it has the same advantages over the prior art as claim 12.

44. Claim 14 of the '243 Patent incorporates the limitations of claim 11, and thus, its approach has the same advantages over the prior art as claim 11.

45. In addition, claim 14 of the '243 Patent recites generating a facial signature from a region of a second (higher-resolution) image corresponding to the location of the head in the reduced-resolution image, and comparing the generated facial signature to a database of facial signatures. In other words, the claim recites finding the location of the head in the first image, and using that location in the second image (which is higher-resolution than the first image) for the generation of a facial signature. That facial signature is compared to reference facial signatures for identification of the audience member. The approach of this claim was not well-understood, routine, or conventional in the prior art. As compared to the prior art, the approach of claim 14 is an improvement because it employs reduced-resolution images where possible (*i.e.,* when locating the head). In turn, as explained above, the advantages of using reduced-resolution images are (1) the ability to accomplish most operations in a poorly-illuminated environment; (2) reduction in computational and storage burdens (due to fewer pixels being processed and stored); and (3) the opportunity to form low-resolution images through the use of contrast-increasing processing.

46. Claim 18 of the '243 Patent (which also incorporates the limitations of claim 16) recites capturing first and second images of a media exposure environment with a camera and analyzing the first image to detect the head of an audience member. The claim further recites reducing the resolution of the first image and determining the orientation of the audience member's head with respect to the camera using the reduced-resolution image. Determining the orientation of an audience member's head based on a reduced-resolution image was not well-understood, routine, or conventional as of December 15, 2011 (the filing date of the parent application from which the '243 Patent issued).

47. An advantage over the prior art of determining the orientation of an audience member's head based on a reduced-resolution image, as recited in claim 18, is that a reduced-resolution image can be obtained in a poorly-illuminated environment. This is so because, as explained above, a poorly-illuminated image can undergo processing that reduces its resolution in exchange for an increase in contrast level. A further advantage over the prior art of determining the orientation of an audience member's head based on a reduced-resolution image is that processing such images requires fewer computational and storage resources than determining head orientation based on a higher-resolution image (such as a high-resolution image that would be suitable for facial recognition analysis). This is so, as explained above, because a reduced-resolution image contains fewer pixels from which data must be processed and stored. And finally, an additional benefit of determining the orientation of an audience member's head based on a reduced-resolution image is that low-resolution images may be formed by applying contrast-increasing processing (such as pixel binning) to higher-resolution images, resulting in higher-contrast images. This increased contrast allows improved determination of head orientation because it provides for easier distinction between the head and its surroundings.

48. Claim 19 of the '243 Patent incorporates the limitations of claim 18, and thus, its approach has the same advantages over the prior art as claim 18.

49. In addition, claim 19 of the '243 Patent recites identifying a region corresponding to the head within the first image from which the reduced-resolution image was obtained, analyzing a portion corresponding to the identified region in a second (higher-resolution) image, and determining whether the head matches a known person. In other words, claim 19 recites finding the location of the audience member's head in the reduced-resolution image, and using

that location in a higher-resolution image for an analysis that allows determination of the identity of the audience member. The approach of this claim was not well-understood, routine, or conventional in the prior art. As compared to the prior art, the approach of claim 19 is an improvement because it employs reduced-resolution images where possible (*i.e.*, when locating the head). In turn, as explained above, the advantages of using reduced-resolution images are (1) the ability to accomplish most operations in a poorly-illuminated environment; (2) reduction in computational and storage burdens (due to fewer pixels being processed and stored); and (3) increased contrast resulting from the ability to form low-resolution images by applying contrast-increasing processing to higher-resolution images (resulting in improved head location and orientation determination).

50. Claim 20 of the '243 Patent includes the limitations of claim 19, and thus, it has the same advantages over the prior art as claim 19.

51. I declare under penalty of perjury under the laws of the United States of America that the foregoing is true and correct.

Executed on: October 12, 2022

Pierre Moulin

Pierre Moulin

EXHIBIT A

PIERRE MOULIN

University of Illinois
Coordinated Science Laboratory
1308 W. Main St.
Urbana, IL 61801

tel (217) 244–8366
fax (217) 244–8371
Email: *moulin@ifp.uiuc.edu*
WWW: *www.ifp.uiuc.edu/˜moulin*

Education

D.Sc. in Electrical Engineering, *Washington University, St Louis*	08/1987–05/1990
M.Sc. in Electrical Engineering, *Washington University, St Louis*	08/1985–05/1986
Ingénieur Civil Electricien, *FacultéPolytechnique de Mons, Belgium*	09/1979–07/1984

Professional Experience

Professor of ECE and Statistics, *University of Illinois*, Urbana, IL	08/2003–present
Associate Professor, *University of Illinois*, Urbana, IL	08/1999–08/2003
Assistant Professor, *University of Illinois*, Urbana, IL	01/1996–08/1999
Research Scientist, *Bell Communications Research*, Morristown, NJ	10/1990–12/1995
Research Assistant, *Washington University*, St Louis, MO	10/1987–05/1990
Research Engineer, *FacultéPolytechnique de Mons, Belgium*	10/1984–06/1985
Summer Intern, *Standard Telecommunications Laboratories Ltd.*, U.K.	07/1983–08/1983

Professional Activities

Senior Editorial Board member of *IEEE Journal of Selected Topics in Signal Processing*, 2019—present
Editorial Board member for *Proceedings of IEEE*, 2007-2012
Co-Founder and Editor in Chief for *IEEE Transactions on Information Forensics and Security*, 2005-2008
Area Editor for *IEEE Transactions on Image Processing*, 2002-2006.
Guest Editor for *IEEE Transactions on Signal Processing* supplements on secure media, 2004-2005.
Guest Associate Editor for *IEEE Transactions on Signal Processing*'s special issue on Data Hiding, Apr. 2003.
Guest Associate Editor for *IEEE Transactions on Information Theory*'s special issue on Information-theoretic imaging, Aug. 2000
Associate Editor for *IEEE Transactions on Image Processing*, 1999—2002
Associate Editor for *IEEE Transactions on Information Theory*, 1996—1998

Member, IEEE Image and Multidim. Signal Processing Technical Committee, 1998-2003

Co-chair, Allerton Conference 2007, 2008

Member, Technical Program Committee of miscelleanous major IEEE conferences

Invited organizer, NSF Workshop on Signal Authentication, Orlando, FL, 2002.

Organizer, special sessions on watermarking at Allerton (10/01, 02, 03) and on information-theoretic imaging at Asilomar (11/02)

Invited lectures at MIT, Stanford, Berkeley, Carnegie-Mellon, Harvard, Illinois, Purdue, Michigan, UC San Diego, UC Santa Barbara, Maryland, National U. of Singapore, Nanyang Tech. U. (Singapore), Chinese U. Hong Kong, Hong Kong U. of Science and Technology, Hong Kong Polytechnic, National Taiwan U., Academia Sinica, National Chiao Tung U. (Taiwan), KAIST (Korea), INRIA Sophia-Antipolis (France), Delft (Netherlands), Louvain (Belgium), Tech. U. Munich (Germany), Bell Labs, IBM Research, Microsoft Research, KLA-Tencor, Qualcomm.

Keynote/Plenary talks at Int. Workshop on Digital Watermarking (IWDW), Seoul, Korea, 2002; Int. Symp. on Image and Signal Processing and Analysis (ISPA), Rome, 2003; Wav-ila Challenge, Barcelona, Spain, 2005; IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), Toulouse, France, 2006; Workshop on Multimedia Content Representation, Classification and Security (MRCS), Istanbul, Turkey, 2006; 3rd Int. Conf. on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), Kaohsiung, Taiwan, 2007; IEEE Int. Conf. on Image Processing (ICIP), Brussels, Belgium, 2011; EUROCON, Zagreb, Croatia, 2013; ICSIPA, Melaka, Malaysia, 2013; VCIP, Singapore, 2015; ISPACS, Taipei, 2019.

Tutorial lecturer at ICIP'01, ICASSP'02, ICIP'04, and ISIT'06.

Co–Chair, IEEE Info Theory Workshop in Detection, Classif. & Imaging (Santa Fe, 2/99); Allerton Conf. (Monticello, IL, 9/06, 9/07); IEEE Workshop on Information Forensics and Security (Tenerife, Spain, 11/12); Tech. Program co-chair, IEEE Symp. on Information Theory (Hong Kong, 6/15).

Member, *IEEE Int. Conf. on Image Processing* Organizing Committee (Chicago 1998; Technical Program Co-chair, Brussels 2011)

Reviewer and panelist for National Science Foundation, reviewer for IEEE Trans. on SP, IP, IT, COM, NN and JOSA, CVGIP, JVCIP, Addison–Wesley, IEEE Press, SIAM, and Annals of Statistics.

Consultant and expert witness

Founding Director of *Center for Information Forensics*, UIUC.

Honors and Awards

Ronald W. Pratt Faculty Outstanding Teaching Award, 2018
IEEE Information Theory Society Board of Governors, 2016—2018
IEEE Signal Processing Society Distinguished Lecturer, 2012–2013
UIUC Sony Faculty Scholar, 2005—2007
IEEE Signal Processing Society Board of Governors, 2005—2007
IEEE Fellow, 2003
Beckman Associate in UIUC Center for Advanced Study, 2003
Co-author, IEEE Signal Processing Society 2002 Young Author Best Paper Award
IEEE Signal Processing Society 1997 Best Paper Award
NSF Career award, 1998-2001
UIUC Incomplete List of Teachers Rated as Excellent, 1996, 1999, 2000, 2005, 2007, 2009, 2011, 2012, 2016, 2019
Rotary International Scholar, 1985-1986
A. Dosin Award Recipient, 1984
Co-winner 1989 US Amateur Team Chess Championship

University Service

Chair, ECE Faculty Search Committee, 2019—Present
ECE Promotions and Tenure Committee, 2009-Present
Chair, ECE Signal Processing and Circuits Area, 2007-2017
ECE Faculty Search Committee, 2006—2012 and 2015—Present
ECE Graduate Committee, 1997—2008 and 2011—2014
ECE Representative to Research Board, 2007-2008
Elected Member, UIUC Faculty Senate, 2004-2005
Member, Information Trust Institute educational committee, 2004 - 2006
Elected Member, ECE Faculty Advisory Committee, 2002-2003
Member, Computer Resources and Education Committee, 1999—2002
Chair, ECE Graduate Seminar Committee, 1998—1999
Member, ECE Graduate Seminar Committee, 1996—2003

Ph.D. Graduates

	Name	Year	Placement
1.	Juan (Julia) Liu	2001	Research Scientist, PARC, Palo Alto, CA
2.	Prakash Ishwar	2002	Prof., Boston U.
3.	Kivanc Mihcak	2002	Assoc Prof., Bogazici U., Istanbul, Turkey
4.	Shawn C. Herman	2002	Numerica Inc., Fort Collins, CO.
5.	Tie Liu	2006	Assoc. Prof., Texas AM
6.	Ying (Grace) Wang	2006	Flarion Qualcomm, NJ
7.	Negar Kiyavash (co-advisor)	2006	Assoc. Prof., UIUC
8.	Maha El Choubassi	2008	Intel Research, Santa Clara, CA
9.	Shankar Sadasivam	2011	Qualcomm, San Diego, CA
10.	Yen-Wei Huang	2013	Google
11.	Honghai Yu	2015	A*STAR, Singapore
12.	Patrick Johnstone	2017	Postdoc, Rutgers U.
13.	Amish Goel	2022	

Postdocs

	Name	Year	Placement
1.	Aaron Lanterman	1998-2001	Prof, Georgia Tech
2.	Sviatoslav Voloshnykovskiy	1999	Assoc. Prof., U. of Geneva
3.	Jong C. Ye	1999-2001	Prof, KAIST, Korea
4.	Natalia Schmid	2001	Assoc. Prof., U. of West Virginia
5.	Bing Bing Ni	2010-2015	Asst. Prof., Shanghai Jiaotong U., China
6.	Jiwen Lu	2012-2015	Asst. Prof., Tsinghua U., Beijing, China
7.	Gang Wang	2012-2015	Assoc. Prof., Nanyang Tech. U., Singapore
8.	Zhang Le	2016-	ADSC Singapore
9.	Jagan Varadarajan	2013 - 2016	ADSC Singapore

Funding NSF, DARPA, ARO, ONR, NATO, A*STAR, Microsoft, HP Labs, Xerox, KLA-Tencor

Sabbatical Leaves

National Taiwan University, 2019

Chinese University of Hong Kong, 2010 and 2014

MIT, 2005

Microsoft Research, 2001

Course Development

1. ECE418, Introduction to Image and Video Processing, 1996-1998
2. ECE544 Wavelets, 1998
3. ECE544 Statistical Image and Video Processing, 2006-2008
4. ECE544 Statistical Learning, 2012, 2015, 2016
5. ECE598PM (now ECE566) Computational Inference and Learning, 2014-2017
6. ECE398BD (now ECE314) Making Sense of Big Data, 2014.

Other Teaching

Introduction to Image and Video Processing (ECE418), Digital Signal Processing II (ECE551), Information Theory (ECE563), Random Processes (ECE534), Detection and Estimation Theory (ECE561)

Journal Papers

A. Goel and P. Moulin, “Fast Locally Optimal Detection of Targeted Universal Adversarial Perturbations,” *IEEE Trans. on Information Forensics and Security*, Vol. 17, pp. 1757— 1770, 2022.

P. R. Johnstone and P. Moulin, “Faster Subgradient Methods for Functions with Hölderian Growth,” *Mathematical Programming*, Vol. 180, pp. 418—450, 2020.

R. Trabelsi, J. Varadarajan, L. Zhang, I. Jabri, Y. Pei, F. Smach, A. Bouallegue, and P. Moulin, ”Understanding the Dynamics of Social Interactions: A Multi-Modal Multi-View Approach” ACM Transactions on Multimedia Computing, Communications, and Applications, Feb. 2019.

R. W. Yeung, C. Chen, Q. Chen, and P.Moulin, On Characterizations of Markov Random Fields and Subfields, *IEEE Trans. Information Theory*, Aug. 2018.

P. Moulin, “The Log-Volume of Optimal Codes on Memoryless Channels, Asymptotically Within a Few Nats,” *IEEE Trans. Information Theory*, Vol. 63, pp. 2278—2313, April 2017.

P. R. Johnstone and P. Moulin, “Local and global convergence of a general inertial proximal splitting scheme for minimizing composite functions,” *Computational Optimization and Applications*, Feb. 2017.

J. Lu, G. Wang and P. Moulin, “Localized Multi-Feature Metric Learning for Image Set Recognition,” *IEEE Trans. Circ. Syst. Video Technology*, Vol. 26, No. 3, pp. 529—540, March 2016.

B. Ni, V. R. Paramathayalan, T. Li, and P. Moulin, “Multiple Granularity Modeling: A Coarse-to-Fine Framework for Fine-grained Action Analysis,” *International Journal on Computer Vision* (IJCV), March 2016.

H. Yu and P. Moulin, “SNR Maximization Hashing,” *IEEE Trans. Information Forensics and Security*, Vol.10, No. 9, pp. 1927—1938, Sep. 2015.

B. Ni, G. Wang, and P. Moulin, “Order Preserving Sparse Coding” *IEEE Trans. on Pattern Recognition and Machine Intelligence*, Aug. 2015, Vol. 37, No. 8, pp. 1615—1628.

H. Yu and P. Moulin, “Regularized Adaboost Learning for Identification of Time-Varying Content,” *IEEE Trans. Information Forensics and Security*, Vol. 9, No. 10, pp. 1606— 1616, Oct. 2014.

Y.-W. Huang and P. Moulin, “On the Fingerprinting Capacity Games for Arbitrary Alphabets and Their Asymptotics,” *IEEE Trans. Information Forensics and Security*, Vol. 9, No. 9, pp. 1477—1490, Sep. 2014.

B. Ni, P. Moulin and S. Yan, “Pose Adaptive Motion Feature Pooling for Human Action Analysis,” *International Journal on Computer Vision* (IJCV), July 2014.

J. Lu, G. Wang and P. Moulin, “Human Identity and Gender Recognition from Gait Sequences with Arbitrary Walking Directions,” *IEEE Trans. on Information Forensics and Security*, Vol. 9, No. 1, pp. 41—51, Jan. 2014.

B. Ni, Y. Pei, S. Yan, and P. Moulin, “Multi-Level Depth and Image Fusion for Human Activity Detection,” *IEEE Transactions on System, Man and Cybernetics*, Part B (T-SMC-B), Vol. 43, No. 5, pp.1383—1394, 2012.

Y.-W. Huang and P. Moulin, “On the Saddle-point Solution and the Large-Coalition Be-

havior of Fingerprinting Games," *IEEE Trans. on Information Forensics and Security*, Vol. 7, No. 1, pp. 160—175, 2012.

S. Sadasivam, P. Moulin and T. P. Coleman, "A Message Passing Approach to Combating Desynchronization Attacks," *IEEE Trans. on Information Forensics and Security*, Vol. 6, pp. 894–905, 2011.

J.-F. Jourdas and P. Moulin, "High-Rate Random-Like Fingerprinting Codes with Linear Decoding Complexity," *IEEE Transactions on Information Forensics and Security*, Vol. 4, No. 4, Dec. 2009.

N. Kiyavash, P. Moulin and T. Kalker, "Regular Simplex Fingerprints and Their Optimality Properties," *IEEE Transactions on Information Forensics and Security*, Vol. 4, No. 3, pp. 318—329, Sep. 2009.

N. Kiyavash and P. Moulin, "Performance of Orthogonal Fingerprints Under Worst-Case Noise Attacks," *IEEE Transactions on Information Forensics and Security*, Vol. 4, No. 3, pp. 293—301, Sep. 2009.

M. El Choubassi and P. Moulin, "On Reliability and Security of Randomized Detectors Against Sensitivity Analysis Attacks," *IEEE Trans. Information Forensics and Security*, Vol. 4, No. 3, pp. 273—283, Sep. 2009.

S. Sadasivam and P. Moulin, "On Estimation Accuracy of Desynchronization Attack Channel Parameters," *IEEE Transactions on Information Forensics and Security*, Vol. 4, No. 3, pp. 284—292, Sep. 2009.

P. Moulin, "A Neyman-Pearson Approach to Universal Erasure and List Decoding," *IEEE Trans. Information Theory*, Oct. 2009.

Y. Wang and P. Moulin, "Perfectly Secure Steganography: Capacity, Error Exponents, and Code Constructions," *IEEE Transactions on Information Theory*, Special Issue on Security, Vol. 54, No. 6, pp. 2706—2722, June 2008.

M. El Choubassi and P. Moulin, "Noniterative Algorithms for Sensitivity Analysis Attacks," *IEEE Trans. Information Forensics and Security*, Vol. 2, No.3, pp. 113—126, June 2007.

P. Moulin and Y. Wang, "Capacity and Random-Coding Exponents for Channel Coding with Side Information," *IEEE Trans. on Information Theory*, Vol. 53, No. 4, pp. 1326— 1347, Apr. 2007.

Y. Wang and P. Moulin, "Optimized Feature Extraction for Learning-Based Image Steganalysis," *IEEE Trans. Information Forensics and Security*, Vol. 2, No. 1, pp. 31—45, March 2007.

P. Moulin, "Signal Transmission with Known-Interference Cancellation," *IEEE Signal Processing Magazine*, Lecture Notes, Vol. 24, No. 1, pp. 134—136, Jan. 2007.

J. C. Ye, P. Moulin and Y. Bresler, "Asymptotic Global Confidence Regions for Parametric 3-D Shape Estimation," *IEEE Trans. Image Processing*, Vol. 15, No. 10, pp. 2904—2919, Oct. 2006.

P. Moulin and A. K. Goteti, "Block QIM Watermarking Games," *IEEE Trans. Information Forensics and Security*, Vol. 1, No. 3, pp. 293–310, Sep. 2006.

S. Jana and P. Moulin, "Optimality of KLT for Encoding Gaussian Vector-Scale Mixtures: Application to Reconstruction, Estimation and Classification," *IEEE Trans. on Informa-*

tion Theory, Vol. 52, No. 9, pp. 4049—4067, Sep. 2006.

T. Liu, P. Moulin and R. Koetter, “On Error Exponents of Modulo Lattice Additive Noise Channels,” *IEEE Trans. on Information Theory*, Vol. 52, No. 2, pp. 454-471, Feb. 2006.

P. Moulin and R. Koetter, “Data-Hiding Codes,” invited tutorial paper, *Proc. IEEE*, Vol. 93, No. 12, pp. 2083–2127, Dec. 2005.

P. Ishwar and P. Moulin, “On the Existence and Characterization of the Maxent Distribution Under General Moment Inequality Constraints,” *IEEE Trans. on Information Theory*, Vol. 51, No. 9, pp. 3322–3333, Sep. 2005.

J. L. Cannons and P. Moulin, “Design and Statistical Analysis of a Hash-Aided Image Watermarking System,” *IEEE Trans. on Image Processing*, Vol. 13, No. 10, pp. 1393—1408, Oct. 2004.

P. Moulin and M. K. Mıhçak, “The Parallel-Gaussian Watermarking Game,” *IEEE Trans. on Information Theory*, Vol. 50, No. 2, pp. 272-289, Feb. 2004.

P. Moulin, “Comments on ”Why Watermarking is Nonsense,” IEEE Signal Processing Magazine, Vol. 20, No. 6, pp. 57-59, Nov. 2003.

P. Moulin and A. Ivanovié, “The Zero-Rate Spread-Spectrum Watermarking Game,” *IEEE Transactions on Signal Processing*, Vol. 51, No. 4, pp. 1098-1117, Apr. 2003.

P. Moulin and J. A. O’Sullivan, “Information-Theoretic Analysis of Information Hiding,” *IEEE Trans. on Information Theory*, Vol. 49, No. 3, pp. 563-593, March 2003.

P. Ishwar and P. Moulin, “On the Equivalence Between Set-Theoretic and Maxent MAP Estimation,” *IEEE Trans. on Signal Processing*, Vol. 51, No. 3, pp. 698-713, March 2003.

J. C. Ye, Y. Bresler and P. Moulin, “Cramer-Rao Bounds for Parametric Shape Estimation in Inverse Problems,” *IEEE Trans. on Image Processing*, Vol. 12, No. 1, pp. 71-84, Jan. 2003.

J. C. Ye, Y. Bresler and P. Moulin, “A Self-Referencing Level-Set Method for Image Reconstruction from Sparse Fourier Samples,” invited paper, *Int. J. of Computer Vision*, special issue on level-set methods, Dec. 2002.

P. Moulin and M. K. Mıhçak, “A Framework for Evaluating the Data-Hiding Capacity of Image Sources,” *IEEE Trans. on Image Processing*, Vol. 11, No. 9, pp. 1029–1042, Sep. 2002.

A. Jain, P. Moulin, M. I. Miller and K. Ramchandran, “Information-Theoretic Bounds on Target Recognition Performance Based on Degraded Image Data,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 9, pp. 1153—1166, Sep. 2002.

J. Liu and P. Moulin, “Information-Theoretic Analysis of Interscale and Intrascale Dependencies Between Image Wavelet Coefficients,” *IEEE Trans. on Image Processing*, Vol. 10, No. 10, pp. 1647—1658, Nov. 2001.

P. Moulin, invited discussion of “Regularization of Wavelet Approximations,” by A. An-toniadis and J. Fan, *Journal of the American Statistical Association*, Vol. 96, No. 455, pp. 959—960, Sep. 2001.

P. Moulin, “The Role of Information Theory in Watermarking and Its Application to Image Watermarking,” invited paper, *Signal Processing*, Vol. 81, No. 6, pp. 1121—1139, June 2001.

J. Liu and P. Moulin, "Complexity-Regularized Image Denoising," *IEEE Trans. on Image Processing*, Vol. 10, No. 6, pp. 841—851, June 2001.

J. C. Ye, Y. Bresler and P. Moulin, "Cramer-Rao Bounds for Parametric Boundaries of Targets in Inverse Scattering Problems," *IEEE Trans. on Antennas and Propagation*, May 2001.

M. K. Mıhçak, P. Moulin, M. Anitescu, and K. Ramchandran, "Rate–Distortion–Optimal Subband Coding Without Perfect Reconstruction Constraints," *IEEE Trans. on Signal Processing*, Vol. 49, No. 3, pp. 542—557, Mar. 2001.

P. Ishwar and P. Moulin, "On Spatial Adaptation of Motion Field Smoothness in Video Coding," *IEEE Trans. Circ. Syst. Video Tech.*, Vol. 10, No. 6, pp. 980–989, Sep. 2000.

P. Moulin and J. Liu, "Statistical Imaging and Complexity Regularization," *IEEE Trans. on Information Theory*, Special issue on information-theoretic imaging, Vol. 46, No. 5, pp. 1762—1777, Aug. 2000.

J. C. Ye, Y. Bresler and P. Moulin, "Asymptotic Global Confidence Regions in Parametric Shape Estimation Problems," *IEEE Trans. on Information Theory*, Special issue on information-theoretic imaging, Vol. 46, No. 5, pp. 1881–1895, Aug. 2000.

P. Moulin, M. Anitescu, and K. Ramchandran, "Theory of Rate–Distortion–Optimal, Constrained Filter Banks — Application to FIR and IIR Biorthogonal Designs," *IEEE Trans. on Signal Processing*, Vol. 48, No. 4, pp. 1120–1132, April 2000.

M. K. Mıhçak, I. Kozintsev, K. Ramchandran and P. Moulin, "Low-Complexity Image Denoising Based on Statistical Modeling of Wavelet Coefficients," *IEEE Signal Processing Letters*, Vol. 6, No. 12, pp. 300—303, Dec. 1999.

R. Krishnamurthy, J. W. Woods and P. Moulin, "Frame Interpolation and Bidirectional Prediction of Video Using Compactly-Encoded Optical Flow Fields and Label Fields," *IEEE Trans. Circ. Syst. Video Tech.*, Vol. 9, No. 5, pp. 713—726, Aug. 1999.

P. Moulin and J. Liu, "Analysis of Multiresolution Image Denoising Schemes Using Generalized–Gaussian and Complexity Priors," *IEEE Trans. on Info. Theory*, Special Issue on Multi-scale Analysis, Vol. 45, No. 3, pp. 909-919, Apr. 1999.

V. Pavlovic, P. Moulin and K. Ramchandran, "An Integrated Framework for Adaptive Subband Image Coding," *IEEE Trans. on Signal Processing*, Vol. 47, No. 4, pp. 10241038, Apr. 1999.

P. Moulin and M. K. Mıhçak, "Theory and Design of Signal-Adapted FIR Paraunitary Filter Banks," *IEEE Trans. on Signal Processing*, Special Issue on Wavelets and Filter Banks, Vol. 46, No. 4, pp. 920—929, Apr. 1998.

P. Moulin, R. Krishnamurthy, "Multiscale Modeling and Estimation of Motion Fields for Video Coding, *IEEE Trans. on Image Processing*, Vol. 6, No. 12, pp. 1606—1620, Dec. 1997.

P. Moulin, M. Anitescu, K.O. Kortanek and F. Potra, "The Role of Linear Semi–Infinite Programming in Signal–Adapted QMF Bank Design," *IEEE Trans. on Signal Processing*, Vol. 45, No. 9, pp. 2160—2174, Sep. 1997.

P. Moulin, "A Multiscale Relaxation Technique for SNR Maximization in Nonorthogonal Subband Coding," *IEEE Trans. on Image Processing*, Vol. 4, No. 9, pp. 1269—1281, Sep. 1995.

P. Moulin, A.T. Ogielski, G. Lilienfeld and J.W. Woods: ”Video Signal Processing and Coding on Data–Parallel Computers,” *Digital Signal Processing: A Review Journal*, Vol. 5, No. 2, pp. 118—129, Apr. 1995.

P. Moulin, invited discussion of “Wavelet Shrinkage: Asymptotia?” by D. Donoho and I. Johnstone, *Journal of the Royal Statistical Society B*, Vol. 57, No. 1, 1995.

P. Moulin, ”Wavelet Thresholding Techniques for Power Spectrum Estimation,” *IEEE Trans. on Signal Processing*, Vol. 42, No. 11, pp. 3126–3136, Nov. 1994.

P. Moulin, ”A Wavelet Regularization Method for Diffuse Radar–Target Imaging and Speckle–Noise Reduction,” *Journal of Mathematical Imaging and Vision*, Special Issue on Wavelets, Vol. 3, No. 1, pp. 123–134, 1993.

J. A. O’Sullivan, P. Moulin, and D. L. Snyder, ”An Application of Splines to Maximum Likelihood Radar Imaging,” *International Journal of Imaging Systems and Technology*, Vol. 4, pp. 256–264, 1993.

P. Moulin, J. A. O’Sullivan, and D. L. Snyder, ”A Method of Sieves for Multiresolution Spectrum Estimation and Radar Imaging,” *IEEE Trans. on Information Theory*, Special Issue on Wavelets and Multiresolution Analysis, pp. 801–813, Mar. 1992.

Book

P. Moulin and V. Veeravalli, *Statistical Inference for Engineers and Data Scientists*, Cambridge University Press, 2019.

Book Chapters

B. Ni, G. Wang, and P. Moulin, “RGBD-HuDaAct: A color-depth video database for human daily activity recognition,” Consumer Depth Cameras for Computer Vision: Advances in Computer Vision and Pattern Recognition, pp 193-208, Springer-Verlag, 2013.

P. Moulin, “Information-Hiding Games,” Springer-Verlag Lecture Notes in Computer Sciences, Vol. 2613, 2003.

P. Moulin, ”Multiscale Image Decompositions and Wavelets,” Handbook of Image and Video Processing. A. C. Bovik, Ed., Academic Press, 2000.

K.O. Kortanek and P. Moulin, ”Using Semi–Infinite Programming in Orthogonal Wavelet Filter Design,” *Semi–Infinite Programming and Its Applications*, Kluwer Academic series on Nonconvex Optimization and its Applications, 1998.

P. Moulin, J. A. O’Sullivan, and D. L. Snyder, ”A Sieve–Constrained Maximum–Likelihood Method for Target Imaging,” in *Radar and Sonar II*, Springer–Verlag, The IMA Volumes in Mathematics and Its Applications, Vol. 39, Eds. F. Grunbaum et al., pp. 95–122, 1992.

P. Moulin, ”Adaptive Multiresolution Image Restoration and Compression,” in *IEEE Case Studies in Medical Instrument Design*, Eds. H. Troy Nagle and W.J. Tomkins, pp. 247–254, IEEE, New York, 1992.

Conference Papers

A. Rakin, Y. Wang, S. Aaron, T. Koike-Akino, P. Moulin and K. Pearsons, “Towards Universal Adversarial Examples and Defenses,” Proc. IEEE Information Theory Workshop, Kanazawa, Japan, Oct. 2021.

Ye Wang, Shuchin Aeron, Adnan Rakin, Toshiaki Koike-Akino and P. Moulin, “Robust Machine Learning via Privacy/Rate-Distortion Theory,” Proc. IEEE Int. Symp. On Information Theory, Melbourne, Australia, July 2021.

A. Goel and P. Moulin, "Locally Optimal Detection of Stochastic Targeted Universal Adversarial Perturbations," Proc. ICASSP, Toronto, Canada, May 2021.

F. Zhuang and P. Moulin, "Deep Semi-Supervised Metric Learning Via Identification of Manifold Memberships" Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, Canada, May 2021.

T. Jayashankar, P. Moulin, and J. Le Roux, "Detecting Audio Attacks on ASR Systems with Dropout Uncertainty," Proc. Interspeech, Oct. 2020.

F. Zhang and P. Moulin, "A New Variational Method for Deep Supervised Image Hashing," Proc. ICASSP, Barcelona, Spain, May 2020.

P. Moulin, "Game-theoretic design of universal adversarial perturbation detectors," presented at *Information Theory and Applications*, San Diego, CA, Feb. 2020.

T. Jayashankar, P. Moulin, T. Blu, C. Gilliam, LAP-Based Video Frame Interpolation, *Proc. ICIP 2019*, Taipei, Taiwan, Sep. 2019.

P. Moulin and A. Goel, " Random Ensemble of Locally Optimum Detectors for Detection of Adversarial Inputs," *IEEE GlobalSIP Conf., Signal Processing for Adversarial Machine Learning*, Anaheim, CA, Nov. 2018.

P. Moulin, "Gaussian Mixture Models for Detection of Adversarial Inputs," *Information Theory and Applications*, San Diego, CA, Feb. 2018.

P. Moulin and A. Goel, " Locally Optimal Detection of Adversarial Inputs to Image Classifiers," *Proc. ICME*, Hong Kong, July 2017.

R. W. Yeung, C. Chen, Q. Chen, and P.Moulin, Information-Theoretic Characterizations of Markov Random Field and Subfield, *Proc. ISIT*, Aachen, Germany, June 2017.

P. Moulin, "Source Coding with Distortion Profile Constraints", *Proc. ISIT*, Aachen, Germany, June 2017.

P. Moulin,"Lower Bounds on Rate of Fixed-Length Source Codes under Average- and f-Fidelity Constraints," *Proc. ISIT*, Aachen, Germany, June 2017.

Zhang Le, J. Varadarajan, P. N. Suganthan, N.Ahuja, and P. Moulin, "Robust Visual Tracking Using Incremental Oblique Random Forest," *CVPR*, June 2017.

J. Varadarajan, R. Subramanian, N. Ahuja, P. Moulin, J.-M. Odobez, "Active Online Anomaly Detection Using Dirichlet Process Mixture Model and Gaussian Process Classification," *Proc. Winter Conference on Applications of Computer Vision*, Santa Rosa, CA, March 2017.

P. R. Johnstone and P. Moulin, "Convergence Rates of Inertial Splitting Schemes for Nonconvex Composite Optimization," *Proc. ICASSP*, New Orleans, LA, March 2017.

P. Moulin, "Refined asymptotic analysis of source coding under excess-distortion constraint," *ITA*, San Diego, CA, Feb. 2017.

P. Moulin, "Asymptotically Achievable Error Probabilities for Multiple Hypothesis Testing," *Proc. IEEE Int. Symp. Information Theory*, Barcelona, Spain, July 2016.

P. Moulin, "Strong Large Deviations for Composite Multiple Hypothesis Testing," *ITA*, San Diego, CA, Feb. 2016.

H. Yu and P. Moulin, "Multi-Feature Hashing Based on SNR Maximization," *Proc. ICIP*, Quebec City, Canada, Sep. 2015.

T. Blu, P. Moulin, and C. Gilliam, "Approximation Order of the LAP Optical Flow Algorithm," *Proc. ICIP*, Quebec City, Canada, Sep. 2015.

P. Moulin, "Coding Redundancy of Finite-Blocklength Universal Channel Coding," *Proc. IEEE Int. Symp. Information Theory*, Hong Kong, June 2015.

P. Moulin and P. R. Johnstone, "Strong Large Deviations for GLRT and Variants in Composite Hypothesis Testing," *Proc. IEEE Int. Symp. Information Theory*, Hong Kong, June 2015.

J. Lee, M. Raginsky, and P. Moulin, "On MMSE Estimation from Quantized Observations in the Nonasymptotic Regime," *Proc. IEEE Int. Symp. Information Theory*, Hong Kong, June 2015.

B. Ni and P. Moulin, "Motion Part Regularization: Improving Action Recognition via Trajectory Selection," *Computer Vision and Pattern Recognition (CVPR)*, June 2015.

J. Lu, G. Wang, W. Deng, P. Moulin, and J. Zhou, "Multi-Manifold Deep Learning for Image Set Classification," *Proc. CVPR*, June 2015.

V. E. Liong, J. Lu, G. Wang, P. Moulin, and J. Zhou, "Deep Hashing for Compact Binary Codes Learning," *Proc. CVPR*, June 2015.

H. Yu and P. Moulin, "SNR Maximization Hashing," *Proc. ICASSP*, Brisbane, Australia, Apr. 2015.

P. R. Johnstone and P. Moulin, "Convergence of an Inertial Proximal Method for L_1 Regularized Least Squares," *Proc. ICASSP*, Brisbane, Australia, Apr. 2015.

S. D. Chen and P. Moulin, "A Classification Centric Quantizer for Efficient Encoding of Predictive Feature Errors," *Proc. Asilomar Conf.*, Monterrey, CA, Nov. 2014.

J. Lu, G. Wang, W. Deng, and P. Moulin, "Simultaneous feature learning and dictionary learning for image set base face recognition," *ECCV*, Lecture Notes in Computer Science Vol. 8689, pp. 265—280, Zurich, Switzerland, Sep. 2014.

Y. Zhou, B. Ni, S. Yan, P. Moulin, and Q. Tian, "Pipelining Localized Semantic Features for Fine-Grained Action Recognition," *ECCV*, Lecture Notes in Computer Science Vol. 8692, pp. 481—496, Zurich, Switzerland, Sep. 2014.

V. Tan and P. Moulin, "Second- and Higher-Order Asymptotics For Erasure and List Decoding," *Proc. ISIT*, Honolulu, HI, pp. 1887—1891, July 2014.

Y.-W. Huang and P. Moulin, "Strong Large Deviations for Composite Hypothesis Testing," *Proc. ISIT*, Honolulu, HI, pp. 556—560, July 2014.

B. Ni, T. Li, and P. Moulin, "Beta Process Multiple Kernel Learning," *CVPR*, Colombus, OH, pp. 963—970, June 2014.

B. Ni, V. R. Paramathayalan and P. Moulin, "Multiple Granularity Analysis for Fine-Grained Action Detection," *CVPR*, Colombus, OH, pp. 756—763, June 2014.

S. D. Chen and P. Moulin, "A Two-Part Predictive Coder for Multitask Signal Compression," *Proc. ICASSP*, Florence, Italy, pp. 2035—2039, May 2014.

R. Naini and P. Moulin, " Fingerprint Information Maximization for Content Identification," *Proc. ICASSP*, Florence, Italy, pp. 3809—3813, May 2014.

P. Moulin and P. R. Johnstone, "Kullback-Leibler Divergence and the Central Limit Theorem," *Proc. ITA*, San Diego, CA, Feb. 2014.

J. Lu, G. Wang, and P. Moulin, "Image set classification using holistic multiple order statistics features and localized multi-kernel metric learning," *IEEE International Conference on Computer Vision* (ICCV), pp. 329—336, 2013.

B. Ni and P. Moulin, "Manipulation Pattern Discovery: A Nonparametric Bayesian Approach," *ICCV*, 2013.

P. Moulin, "Asymptotic Neyman-Pearson Games for Converse to the Channel Coding Theorem," *Proc. ISIT*, Istanbul, July 2013.

P. R. Johnstone, A. Emad, P. Moulin, and O. Milenkovic, "RFIT: A New Algorithm for Matrix Completion," *Proc. SPARS*, Lausanne, Switzerland, July 2013.

D. Sungatullina, J. Lu, G. Wang, and P. Moulin, "Multiview discriminative learning for age-invariant face recognition," *IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pp. 1—6, Shanghai, China, Apr. 2013.

H. Yu and P. Moulin, "Regularized Adaboost for Content identification," *Proc. ICASSP*, Vancouver, Canada, May 2013.

H. Yu, P. Moulin and S. Roy, "RGB-D Video Content Identification," *Proc. ICASSP*, Vancouver, Canada, May 2013.

D. Sungatullina, J. Lu, G. Wang, and P. Moulin, "Multiview Discriminative Learning for Age-Invariant Face Recognition," *Proc. 10th IEEE International Conference on Automatic Face and Gesture Recognition*, Shanghai, China, Apr. 2013.

P. Moulin, "A New Metaconverse and Outer Region for Finite-Blocklength MACs," *Proc. Information Theory and Applications (ITA) Workshop*, San Diego, CA, Feb. 2013.

A. Singh, N. Ahuja, P. Moulin, "Online learning with kernels: Overcoming the growing sum problem," *Proc. IEEE Machine Learning for Signal Processing (MLSP)*, Santander, Spain, pp. 449-454, Sep. 2012.

B. Ni, M. Xu, J. Tang, S. Yan, and P. Moulin, "Omni-range spatial contexts for visual classification," *Proc. CVPR*, pp. 3514-3521, Providence, RI, June 2012.

B. Ni, P. Moulin, S. Yan, "Order-Preserving Sparse Coding for Sequence Classification," *Proc. ECCV* pp. 173-187, Providence, RI, June 2012.

B. Ni, N. C. Dat, P. Moulin, "RGBD-camera based get-up event detection for hospital fall prevention," *Proc. ICASSP*, pp. 1405-1408, Kyoto, Japan, March 2012.

R. Dubey, B. Ni, and P. Moulin, "A Depth Camera Based Fall Recognition System for the Elderly," *Proc. 9th Int. Conf. Image Analysis and Recognition (ICIAR)*, LNCS Vol. 7325, pp. 106-113, Aveiro, Portugal, June 2012.

P. Moulin, "The log-volume of optimal constant-composition codes for memoryless channels, within O(1) bits," *Proc. ISIT*, pp. 826-830, Cambridge, MA, July 2012.

Y.-W. Huang and P. Moulin, "Finite blocklength coding for multiple access channels," *Proc. ISIT*, pp. 831-835, Cambridge, MA, July 2012.

Y.-W. Huang and P. Moulin, "On fingerprinting capacity games for arbitrary alphabets and their asymptotics," *Proc. ISIT*, pp. 2571-2575, Cambridge, MA, July 2012.

R. Naini and P. Moulin, "Real Adaboost for content identification," *Proc. IEEE Stat. Signal Processing Workshop*, Ann Arbor, MI, Aug. 2012.

R. Naini and P. Moulin, "Model-based decoding metrics for content identification," *Proc. ICASSP*, pp. 1829-1832, Kyoto, Japan, March 2012.

P. Moulin, "The log-volume of optimal codes for memoryless channels, Up to a Few Nats," *Proc. Information Theory and Applications Workshop*, San Diego, CA, Feb. 2012.

B. Ni, G. Wang, and P. Moulin, "RGBD-HuDaAct: A color-depth video database for human daily activity recognition," *Proc. 1st IEEE workshop on Consumer Depth Cameras for Computer Vision*, held in conjunction with *Int. Conf. on Computer Vision (ICCV) Workshop*, pp. 1147-1153, Barcelona, Nov. 2011.

S. Sadasivam and P. Moulin, "A Universal Divergence-Rate Estimator for Steganalysis in Timing Channels," *Proc. 2nd Workshop on Information Security and Forensics*, Seattle, WA, Dec. 2010.

Y.-W. Huang and P. Moulin, "On the Saddle-point Solution and the Large-Coalition Behavior of Fingerprinting Games," *Proc. 2nd Workshop on Information Security and Forensics*, Seattle, WA, Dec. 2010.

R. Naini and P. Moulin, "Interference Management through Mobile Relays in Ad Hoc Networks," *Proc. ASILOMAR Conf.*, Pacific Grove, CA, Nov. 2010.

G. Gigaud and P. Moulin, "A Geometrically Resilient SURF-Based Image Fingerprinting Scheme," *Proc. IEEE Int. Conf. Image Processing* (ICIP), Hong Kong, Sep. 2010.

G. Gigaud and P. Moulin, "Traitor Tracing Aided by Compressed SURF Features," *Proc. Conf. Info. Sci. and Sys.*. Princeton, NJ, March 2010.

P. Moulin, "Statistical Modeling and Analysis of Content Identification and Retrieval," *Proc. Information Theory and Applications Workshop*, San Diego, CA, Feb. 2010.

Y.-W. Huang and P. Moulin, "Capacity-Achieving Fingerprint Decoding," *Proc. 1st Workshop on Information Security and Forensics*, London, UK, Dec. 2009.

S. Sadasivam and P. Moulin, "Combating Desynchronization Attacks on Blind Watermarking Systems: A Message Passing Approach," *Proc. ICIP*, Cairo, Egypt, Nov. 2009.

I. Ezzeddine and P. Moulin, *Proc. IEEE Information Theory Workshop*, Taornima, Italy, Oct. 2009.

S. D. Chen, V. Monga, and P. Moulin, "Meta-classifiers For Multimodal Document Classification," *Proc. IEEE Multimedia Signal Processing Workshop*, Rio de Janeiro, Brazil, Oct. 2009.

Y.-W. Huang and P. Moulin, Saddlepoint Solution of the Fingerprinting Capacity Game under the Marking Assumption, *Proc. Int. Symp. Information Theory*, Seoul, Korea, July 2009.

P. Moulin, "A Strong Converse for the Gel'fand-Pinsker Channel," *Proc. Int. Symp. Information Theory*, Seoul, Korea, July 2009.

P. Moulin, "Optimal Gaussian Fingerprint Decoders" *Proc. ICASSP*, Taipei, Taiwan, Apr. 2009.

P. Moulin and Y. Wang, "Information-Theoretic Analysis of Spherical Fingerprinting," Proc. ITA Workshop, San Diego, CA, Feb. 2009.

M. El Choubassi and P. Moulin, "Joint Estimation-Detection Games for Sensitivity Analysis Attacks," *Proc. SPIE Conf.*, San Jose, CA, Jan. 2009.

P. Moulin, "A Neyman-Pearson Approach to Universal Erasure and List Decoding," *Proc. IEEE Symp. Information Theory*, pp. 61—65, July 2008.

P. Moulin, "Universal Fingerprinting: Capacity and Random-Coding Exponents," *Proc. IEEE Symp. Information Theory*, pp. 220—224, July 2008.

J.-F. Jourdas and P. Moulin, "Towards Optimal Design of Multimedia Fingerprinting Codes," *Proc. ICASSP*, Las Vegas, NV, pp. 1649—1652, April 2008.

J.-F. Jourdas and P. Moulin, "A Low-Rate Fingerprinting Code And Its Application To Blind Image Fingerprinting," *Proc. SPIE*, San Jose, CA, Jan. 2008.

J.-F. Jourdas and P. Moulin, "A High-Rate Fingerprinting Code," *Proc. SPIE*, San Jose, CA, Jan. 2008.

S. Sadasivam, P. Moulin, and R. Koetter, "Graphical Models for Desynchronization-Resilient Watermark Decoding," *Proc. IEEE Int. Conf. on Image Proc.*, San Antonio, TX, pp. V.477—480, Sep. 2007.

M. El Choubassi and P. Moulin, "Sensitivity Analysis Attacks Against Randomized Detectors," *Proc. IEEE Int. Conf. on Image Proc.*, San Antonio, TX, pp. II.129—132, Sep. 2007.

N. Kiyavash and P. Moulin, "Digital Fingerprinting: on Colluders Worst-Case Noise," *IEEE Workshop on Statistical Signal Processing*, Madison, WI, July 2007.

P. Moulin and N. Kiyavash, "Expurgated Gaussian Fingerprinting Codes," *Proc. IEEE Int. Symp. on Information Theory*, Nice, France, June 2007.

P. Moulin and N. Kiyavash, "Performance of Random Fingerprinting Codes Under Arbitrary Nonlinear Attacks," *Proc. ICASSP*, Honolulu, HI, pp. II.157—160, April 2007.

N. Kiyavash and P. Moulin, "Sphere Packing Lower Bound on Fingerprinting Error Probability," *Proc. SPIE*, San Jose, CA, Jan. 2007.

Y. Wang and P. Moulin, "Capacity and Optimal Collusion Attack Channels for Gaussian Fingerprinting Games," *Proc. SPIE*, San Jose, CA, Jan. 2007.

S. Sadasivam and P. Moulin, "Cramer-Rao Bound on Watermark Desynchronization Parameter Accuracy," *Proc. SPIE*, San Jose, CA, Jan. 2007.

L. Perez-Freire, P. Moulin, and F. Perez-Gonzalez, "Security of spread-spectrum-based data hiding," in Security, Steganography, and Watermarking of Multimedia Contents IX, Edward J. Delp III and P. W. Wong, Eds., vol. 6505. San Jose, California, USA: SPIE, January 2007.

P. Moulin, "On the Optimal Structure of Watermark Decoders Under Geometric Attacks," *Proc. IEEE Int. Conf. on Image Proc.*, Atlanta, GA, Oct. 2006.

P. Moulin, "Universal Decoding of Watermarks Under Geometric Attacks" *Proc. IEEE Int. Symp. on Information Theory*, July 2006.

Y. Wang and P. Moulin, "Capacity and Random-Coding Error Exponent for Public Fin-

gerprinting Game," *Proc. IEEE Int. Symp. on Information Theory*, July 2006.

S. Ray, P. Moulin and M. Médard, "On Jamming in the Wideband Regime," *Proc. IEEE Int. Symp. on Information Theory*, July 2006.

S. Ray, P. Moulin and M. Médard, "On Optimal Signaling and Jamming Strategies in Wideband Fading Channels," *Proc. IEEE Conf. on Signal Processing Applications for Wireless Communications (SPAWC)*, June-July 2006.

N. Kiyavash and P. Moulin, "On Optimal Collusion Strategies for Fingerprinting", *Proc. ICASSP*, Toulouse, France, May 2006.

N. Kiyavash and P. Moulin, "A Framework for Optimizing Nonlinear Collusion Attacks on Fingerprinting Systems", *Proc. CISS*, Princeton, NJ, March 2006.

P. Moulin and R. Koetter, "A Framework for Design of Good Watermark Identification Codes," *Proc. SPIE*, San Jose, CA, Jan. 2006.

M. El Choubassi and P. Moulin, "On the fundamental tradeoff between watermark detection performance and robustness against sensitivity analysis attacks," *Proc. SPIE*, San Jose, CA, Jan. 2006.

Y. Wang and P. Moulin, "Statistical modeling and steganalysis of DFT-based image steganography," *Proc. SPIE*, San Jose, CA, Jan. 2006.

N. Kiyavash and P. Moulin, "Regular Simplex Fingerprints and Their Optimality Properties," *Proc. Int. Workshop on Digital Watermarking*, Siena, Italy, Sep. 2005.

P. Moulin and Y. Wang, "Improved QIM Strategies for Gaussian Watermarking," *Proc. Int. Workshop on Digital Watermarking*, Siena, Italy, Sep. 2005.

P. Moulin and A. K. Goteti, "Minmax Strategies for QIM Watermarking Subject to Attacks with Memory," *Proc. IEEE ICIP*, Sep. 2005.

P. Moulin and Y. Wang, "On Achievable Error Exponents for Watermarking," *Proc. SPIE Conf.*, San Jose, CA, Jan. 2005

M. El Choubassi and P. Moulin, "A New Sensitivity Analysis Attack," *Proc. SPIE Conf.*, San Jose, CA, Jan. 2005

A. K. Goteti and P. Moulin, "QIM Watermarking Games," *Proc. ICIP*, Singapore, Oct. 2004.

P. Moulin and A. Briassouli, "A Stochastic QIM Algorithm for Robust, Undetectable Image Watermarking," invited paper, *Proc. ICIP*, Singapore, Oct. 2004.

T. Liu, P. Moulin and R. Koetter, "On Error Exponents of Nested Lattice Codes for the AWGN Channel," *Proc. Information Theory Workshop*, San Antonio, TX, Oct. 2004.

P. Moulin and Y. Wang, "Error Exponents for Channel Coding with Side Information," *Proc. Information Theory Workshop*, San Antonio, TX, Oct. 2004.

P. Moulin, A. K. Goteti and R. Koetter, "Optimal Sparse QIM Codes for Zero-Rate Blind Watermarking," *Proc. ICASSP*, Montreal, Canada, May 2004.

A. K. Goteti and P. Moulin, "Two Private, Perceptual Data-Hiding Games," *Proc. ICASSP*, Montreal, Canada, May 2004.

P. Moulin and Y. Wang, "New Results on Steganographic Capacity," *Proc. Conf. on Info.*

Syst. and Sci., Princeton, NJ, March 2004.

Y. Wang and P. Moulin, “Steganalysis of Block-Structured Stegotext,” *Proc. SPIE* Vol. 5306, San Jose, CA, Jan. 2004.

P. Moulin, “Embedded-Signal Design for Channel Parameter Estimation Part I: Linear Embedding,” *Proc. IEEE Workshop on Statistical Signal Processing*, St Louis, MO, Sep. 2003.

P. Moulin, “Embedded-Signal Design for Channel Parameter Estimation Part II: Quantization Embedding,” *Proc. IEEE Workshop on Statistical Signal Processing*, St Louis, MO, Sep. 2003.

Y. Wang and P. Moulin, “Steganalysis of Block-DCT Steganography,” *Proc. IEEE Workshop on Statistical Signal Processing*, St Louis, MO, Sep. 2003.

J. Liu and P. Moulin, “On the Risk of Translation-Averaged Wavelet Estimators,” *Proc. SPIE*, San Diego, Aug. 2003.

S. Jana and P. Moulin, “Optimal Transform Coders for Classification/Reconstruction with Joint Decoding at the Receiver,” *Proc. IEEE Int. Symp. on Information Theory*, Yokohama, Japan, July 2003.

T. Liu and P. Moulin, “Error Exponents for Zero-Rate Watermarking Game with Squared-Error Constraints,” *Proc. IEEE Int. Symp. on Information Theory*, Yokohama, Japan, July 2003.

T. Liu and P. Moulin, “Error Exponents for One-Bit Watermarking,” *IEEE Proc. ICASSP*, Hong Kong, April 2003.

A. Briassouli and P. Moulin, “Detection-Theoretic Anaysis of Warping Attacks in Spread-Spectrum Watermarking,” *IEEE Proc. ICASSP*, Hong Kong, April 2003.

S. Jana and P. Moulin, “Optimal Transform Coding of Gaussian Mixtures for Joint Classification/Reconstruction,” *Proc. Data Compression Conference*, Snowbird, UT, March 2003.

N. Schmid, Y. Bresler and P. Moulin, “Complexity Regularized Shape Estimation from Noisy Fourier Data,” *Proc. IEEE Int. Conf. on Image Processing*, Rochester, NY, 2002.

J. C. Ye, Y. Bresler and P. Moulin, “Cramer-Rao Bounds for Parametric Shape Estimation in Inverse Problems,” *Proc. IEEE Int. Conf. on Image Processing*, Rochester, NY, 2002.

A. Ivanovié and P. Moulin, “Game-Theoretic Performance Analysis of Image Watermark Detectors,” *Proc. IEEE Int. Conf. on Image Processing*, Rochester, NY, 2002.

M. K. Mıhçak and P. Moulin, “Information-Embedding Codes Matched to Local Gaussian Image Models,” *Proc. IEEE Int. Conf. on Image Processing*, Rochester, NY, 2002.

P. Moulin, A. Briassouli and H. Malvar, “Detection-Theoretic Analysis of Desynchroniza-tion Attacks in Watermarking,” *Proc. DSP’02*, Santorini, Greece, July 2002.

P. Moulin and M. T. Orchard, “Undergraduate Education in Image and Video Processing,” *Proc. Int. Conf. on Ac., Sp., and Signal Processing*, Orlando, FL, May 2002.

S. C. Herman and P. Moulin, “A Particle Filtering Approach to Joint Radar Tracking and Automatic Target Recognition”, *Proc. IEEE Aerospace Conference*, Big Sky, Montana, March 2002.

P. Moulin and A. Briassouli, "The Gaussian Fingerprinting Game," *Proc. CISS'02*, Princeton, NJ, March 2002.

P. Moulin and J. A. O'Sullivan, "Optimal Key Design in Information-Embedding Systems," *Proc. CISS'02*, Princeton, NJ, March 2002.

P. Moulin and A. Ivanović, "The Fisher Information Game for Optimal Design of Synchronization Patterns in Blind Watermarking," *Proc. IEEE Int. Conf. on Image Processing*, Thessaloniki, Greece, Oct. 2001.

P. Moulin and A. Ivanović, "Game-Theoretic Analysis of Watermark Detection," *Proc. IEEE Int. Conf. on Image Processing*, Thessaloniki, Greece, Oct. 2001.

J. C. Ye, Y. Bresler and P. Moulin, "A Self-Referencing Level-Set Method for Image Reconstruction from Sparse Fourier Samples," *Proc. IEEE Int. Conf. on Image Processing*, Thessaloniki, Greece, Oct. 2001.

J. Liu and P. Moulin, "Statistical Image Restoration Based on Adaptive Wavelet Models," *Proc. IEEE Int. Conf. on Image Processing*, Thessaloniki, Greece, Oct. 2001.

J. Liu and P. Moulin, "Approximation-Theoretic Analysis of Translation-Invariant Wavelet Denoising," *Proc. IEEE Int. Conf. on Image Processing*, Thessaloniki, Greece, Oct. 2001.

J. A. O' Sullivan and P. Moulin, "Some Properties of Optimal Information Hiding and Information Attacks," *Proc. 39th Allerton Conf.*, pp. 92-101, Monticello, IL, Oct. 3-5, 2001.

J. C. Ye, Y. Bresler and P. Moulin, "A Self-Referencing Level-Set Method for Image Reconstruction from Sparse Fourier Samples," *Proc. IEEE Int. Worskshop on Variational and Level Set Methods*, Vancouver, B.C., July 2001.

P. Moulin and A. Ivanović, "The Watermark Selection Game," *Proc. Conf. on Info. Syst. and Sci.*, Baltimore, MD, March 2001.

P. Moulin, "The Parallel-Gaussian Watermarking Game," *Proc. Conf. on Info. Syst. and Sci.*, Baltimore, MD, March 2001.

J. Liu and P. Moulin, "Approximation-Theoretic Analysis of Translation-Invariant Wavelet Denoising," invited paper, *Proc. Conf. on Info. Syst. and Sci.*, Baltimore, MD, March 2001.

P. Moulin, M. K. Mıhçak and G.-I. Lin, "An Information-Theoretic Model for Image Watermarking," *IEEE Int. Conf. on Image Processing*, Vancouver, BC, Sep. 2000.

R. Venkatesan, M. Jakubowski, S.-M. William Koon and P. Moulin, "Robust Image Hashing," *IEEE Int. Conf. on Image Processing*, Vancouver, BC, Sep. 2000.

P. Ishwar and P. Moulin, "Shift-Invariant Image Denoising — A Multiple-Domain Maxent MAP Framework," *IEEE Int. Conf. on Image Processing*, Vancouver, BC, Sep. 2000.

J. Liu and P. Moulin, "Analysis of Interscale and Intrascale Dependencies Between Image Wavelet Coefficients," *IEEE Int. Conf. on Image Processing*, Vancouver, BC, Sep. 2000.

S. Jana and P. Moulin, "Optimal Design of Transform Coders and Quantizers for Image Classification," *IEEE Int. Conf. on Image Processing*, Vancouver, BC, Sep. 2000.

M. K. Mıhçak, A. F. Docimo, P. Moulin and K. Ramchandran, "Design and Analysis of a Forward-Adaptive Wavelet Image Coder," *IEEE Int. Conf. on Image Processing*, Vancouver, BC, Sep. 2000.

M. Kesal, M. K. Mıhçak, R. Kötter and P. Moulin, "Iteratively Decodable Codes for Watermarking Applications," *Proc. 2nd Symposium on Turbo Codes and Related Topics*, pp. 113-116, Brest, France, Sep. 2000.

P. Ishwar and P. Moulin, "Image Modeling and Restoration – Information fusion, Set-theoretic methods, and the Maximum Entropy Principle," *Proc. MAXENT'00*, Paris, France, July 2000.

P. Moulin and J. A. O'Sullivan, "Information-Theoretic Analysis of Information Hiding," *Proc. IEEE Int. Symp. on Information Theory*, Sorrento, Italy, June 2000.

P. Moulin and J. Liu, "Statistical Imaging and Complexity Regularization," *Proc. IEEE Int. Symp. on Information Theory*, Sorrento, Italy, June 2000.

P. Moulin and J. A. O'Sullivan, "Information-Theoretic Analysis of Watermarking," *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP'00)*, Istanbul, Turkey, June 2000.

P. Ishwar and P. Moulin, "Fundamental Equivalences between Set-Theoretic and Maximum-Entropy methods in Multiple-Domain Image Restoration," *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP'00)*, Istanbul, Turkey, June 2000.

J. C. Ye, Y. Bresler and P. Moulin, "Global Confidence Regions in Parametric Shape Estimation," *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP'00)*, Istanbul, Turkey, June 2000.

A. Jain, P. Moulin, M. I. Miller and K. Ramchandran, "Information-Theoretic Bounds on Target Recognition Performance," *SPIE's 14th International Symposium on AeroSense 2000*, Orlando, FL, April 2000.

J. C. Ye, Y. Bresler and P. Moulin, "Wavelet Descriptors for Parametric Shape Estimation in Nonlinear Inverse Scattering Problems," *SPIE's International Symposium on AeroSense 2000*, Orlando, FL, April 2000.

P. Ishwar and P. Moulin, "Multiple–Domain Image Modeling and Restoration," *Proc. IEEE Int. Conf. on Image Proc.*, Vol. I, pp. 362—366, Kobe, Japan, 1999.

J. Liu and P. Moulin, "Image Denoising Based on Scale–Space Mixture Modeling of Wavelet Coefficients," *Proc. IEEE Int. Conf. on Image Proc.*, Vol. I, pp. 386—390, Kobe, Japan, 1999.

S. Jana and P. Moulin, "Optimal Design of Transform Coders for Image Classification," *Proc. Conf. on Info. Syst. and Sci.*, Baltimore, MD, March 1999.

A. Jain, P. Moulin, K. Ramchandran and M. I. Miller, "Performance Bounds for ATR based on Compressed Data," *Proc. Army Compression Workshop*, Huntsville, AL, Dec. 1998.

P. Moulin, M. Anitescu, and K. Ramchandran, "Some Fundamental Properties of MMSE Filter Banks," *Proc. IEEE Sig. Proc. Symp. on Time-Frequency and Time-Scale Analysis*, Pittsburgh, PA, Oct. 1998.

P. Moulin and J. Liu, "Analysis of Multiresolution Image Denoising Schemes Using Generalized - Gaussian Priors," *Proc. IEEE Sig. Proc. Symp. on Time-Frequency and Time-Scale Analysis*, Pittsburgh, PA, Oct. 1998.

J. Liu and P. Moulin, "A New Complexity Prior for Multi-Resolution Image Denoising," *Proc. IEEE Sig. Proc. Symp. on Time-Frequency and Time-Scale Analysis*, Pittsburgh, PA, Oct. 1998.

J. Liu and P. Moulin, “Complexity-Regularized Image Restoration,” *Proc. IEEE Int. Conf. on Image Proc. (ICIP’98)*, Chicago, IL, Oct. 1998.

M. K. Mıhçak, K. Ramchandran and P. Moulin, “Adaptive Wavelet Packet Image Coding Using an Estimation–Quantization Framework,” *Proc. IEEE Int. Conf. on Image Proc. (ICIP’98)*, Vol. I, pp. 92—96, Chicago, IL, Oct. 1998.

P. Moulin and M. Miller, “Detection Bounds for Automatic Target Recognition,” *Proc. Int. Symp. on Information Theory (ISIT’98)*. Boston, MA, Aug. 1998.

J. A. O’Sullivan, P. Moulin and M. Ettinger, “Information Theoretic Analysis of Steganography,” *Proc. Int. Symp. on Information Theory (ISIT’98)*, Boston, MA, Aug. 1998.

X. Yang, K. Ramchandran and P. Moulin, “Optimization of High–Energy–Compaction, Nearly–Orthonormal, Linear–Phase Filter Banks,” *Proc. ISCAS’98*, Monterrey, CA, June 1998.

P. Ishwar, K. Ratakonda, P. Moulin and N. Ahuja, “Image Denoising in Multiple Compaction Domains,” invited paper, *Proc. ICASSP’98*, Vol. 3, pp. 1889—1892, Seattle, May 1998.

P. Moulin, M. Anitescu and K. Ramchandran, “Asymptotic Performance of Subband Coders Using Constrained, Signal–Adapted FIR Filter Banks,” *Proc. 32nd Conf. on Info. Syst. and Sci.* (CISS), Princeton, NJ, March 1998.

J. Liu and P. Moulin, “Complexity-Regularized Image Denoising,” *Proc. IEEE Int. Conf. on Image Proc. (ICIP’97)*, pp. II. 370—373, Santa Barbara, CA, Oct. 1997.

P. Ishwar and P. Moulin, “Switched Control-Grid Interpolation for Motion-Compensated Video Coding,” *Proc. IEEE Int. Conf. on Image Proc. (ICIP’97)*, pp. III. 650—653, Santa Barbara, CA, Oct. 1997.

R. Krishnamurthy, J. W. Woods and P. Moulin, “Compactly-Encoded Optical Flow Fields for Bidirectional Prediction and Video Coding,” *Proc. VCIP’97*, Feb. 1997.

P. Moulin, K. Ramchandran and V. Pavlovic, “Transform Image Coding based on Joint Adaptation of Filter Banks and Tree Structures,” *Proc. IEEE Int. Conf. on Image Proc. (ICIP’96)*, Lausanne, Switzerland, Sep. 1996.

R. Krishnamurthy, P. Moulin and J. W. Woods, “Multiscale Motion Estimation for Scalable Video Coding,” *Proc. ICIP’96*, Lausanne, Switzerland, pp. I. 965—968, Sep. 1996.

P. Moulin, “Signal Estimation Using Adapted Tree–Structured Bases and the MDL Principle,” *Proc. IEEE Sig. Proc. Symp. on Time-Frequency and Time-Scale Analysis*, pp. 141—143, Paris, June 1996.

P. Moulin, ”Design of Signal–Adapted FIR Paraunitary Filter Banks,” *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP’96)*, Atlanta, GA, May 1996.

P. Moulin and R. Krishnamurthy, ”Multiscale Motion Fields for Video Coding,” *Proc. 9th IEEE Workshop on Image and Multidim. Signal Processing*, Belize, March 1996.

R. Krishnamurthy, P. Moulin and J. Woods, ”Optical Flow Techniques Applied to Video Coding,” *Proc. ICIP’95*, Washington, D.C., Oct. 1995.

P. Moulin, “Model Selection Criteria and the Orthogonal Series Method for Function Estimation,” *Proc. IEEE Int. Symp. on Information Theory (ISIT’95)*, Whistler, B.C., Sep. 1995.

P. Moulin, ”A New Look at the Signal–Adaptive QMF Bank Design,” *Proc. ICASSP’95*, Detroit, MI, pp. 1312—1315, Apr. 1995.

P. Moulin, ”A Relaxation Technique for Minimizing the L^2 Reconstruction Error in Nonorthogonal Subband Coding,” *Proc. ICIP’94*, Austin, TX, Nov. 1994.

P. Moulin, ”The Use of Wavelets for Spectral Density Estimation with Local Bandwidth Adaptation,” *Proc. ISIT’94*, p. 40, Trondheim, Norway, June 1994. Also invited paper, *Interface’94*, Research Triangle Park, June 1994.

P. Moulin, ”Optimization Techniques for Efficient Image Compression in Nonorthogonal Multiresolution Bases,” *Proc. 1994 NJIT Symp. on Wavelets*, Apr. 1994.

P. Moulin, ”Wavelets as a Regularization Technique for Spectral Density Estimation,” *Proc. SIAM Annual Meeting*, p. 69, Philadelphia, July 1993.

P. Moulin, I. Rhee and P. Caputo, ”Video Signal Processing on SIMD Parallel Computer Architectures,” *Proc. 8th IEEE Workshop on Image and Multidim. Signal Processing*, pp. 246–247, Cannes, France, Sep. 1993.

P. Moulin and A. Loui, ”Hierarchical–Finite–Element Estimation and Parameterization of Optical Flow Field,” *Proc. 8th IEEE Workshop on Image and Multidim. Signal Processing*, pp. 130–131, Cannes, France, Sep. 1993.

P. Moulin and A. Loui: ”Application of a Multiresolution Optical Flow Based Method for Motion Estimation to Video Coding,” *Proc. ISCAS’93*, pp. I.1–4, Chicago, IL, May 1993.

P. Moulin, ”Optimal L^2 Approximation of Images in Nonorthogonal Multiresolution Bases,” *Proc. ICASSP’93*, pp. V.189–192, Minneapolis, MN, Apr. 1993.

P. Moulin, ”Model–Based Multiresolution Restoration of Speckle Images: Application to Radar Imaging,” invited paper, *Proc. IEEE Int. Symp. on Information Theory*, p. 67, San Antonio, TX, Jan. 1993.

P. Moulin, ”Adaptive Multiresolution Image Restoration and Compression,” invited paper, *IEEE EMBS Workshop on Case Studies in Medical Instrument Design*, Orlando, FA, Oct. 91.

P. Moulin, ”Wavelets as a Regularization Technique for Spectral Density Estimation,” *Proc. IEEE Signal Processing Conf. on Time–Frequency and Time–Scale Analysis*, pp. 73–76, Victoria, B.C., Oct. 1992.

P. Moulin, ”An Adaptive Finite–Element Method for Image Representation,” *Proc. Int. Conf. on Pattern Recognition*, pp. III.70–74, The Hague, The Netherlands, Aug. 1992.

P. Moulin, ”An Adaptive Spline Method for Signal Restoration,” *Proc. ICASSP’92*, San Francisco, CA, Mar. 92, pp. IV.169–172.

P. Moulin, J. A. O’Sullivan, and D. L. Snyder, ”A Method of Sieves for Estimating the Spectral Density,” *Proc. IEEE Int. Symp. on Information Theory*, p. 67, Budapest, Hungary, June 91.

J. A. O’Sullivan, P. Moulin, D. L. Snyder, and S. P. Jacobs: ”Computational Considerations for Maximum–Likelihood Radar Imaging,” *Proc. 24th CISS Conf.*, Princeton, NJ, March 1990.

P. Moulin, D. L. Snyder, and J. A. O’Sullivan: ”A Method of Sieves for Regularizing Maximum–Likelihood Spectral Estimates,” *Proc. IEEE Int. Symp. on Information The-*

ory, p. 112, San Diego, CA, Jan. 1990.

P. Moulin, D. L. Snyder, and J. A. O'Sullivan: "A Sieve–Constrained Maximum–Likelihood Estimator for the Spectrum of a Gaussian Process," *Proc. 28th Allerton Conf.*, Urbana– Champaign, IL, Sept. 1989.

J. A. O'Sullivan, D. L. Snyder, and P. Moulin, "The Role of Spectrum Estimation in Forming High–Resolution Radar Images," *Proc. ICASSP 1989*, Glasgow, U.K., May 1989.

P. Moulin, J. A. O'Sullivan, and D. L. Snyder, "Maximum–Likelihood Spectrum Estimation of Periodic Processes from Noisy Data," *Proc. CISS Conf.*, Johns Hopkins University, Baltimore, MD, March 1989.

J. A. O'Sullivan, P. Moulin, and D. L. Snyder, "Cramer–Rao Bounds for Constrained Spectrum Estimation with Application to a Problem in Radar Imaging," *Proc. 27th Allerton Conf.*, Urbana–Champaign, IL, Sept. 1988.

R. Boite and P. Moulin, "Comparaison du rapport signal à bruit pour une représentation numérique en virgule fixe et en virgule flottante," *Proc. IASTED Symp.*, Paris, France, June 1985.

EXHIBIT B

Materials Considered

Stephen Shankland,

"Samsung S21 Ultra uses pixel binning to make your photos better"

CNET, Jan. 2021.

www.cnet.com/tech/mobile/samsung-s21-ultra-pixel-binning-adapts-photos-for-detail-or-the-dark/

Rabia Jafri and Hamid R. Arabnia,

"A Survey of Face Recognition Techniques"

Journal of Information Processing Systems, Vol.5, No.2, June 2009

https://www.cosy.sbg.ac.at/~uhl/face_recognition.pdf

Cha Zhang and Zhengyou Zhang,

"A Survey of Recent Advances in Face Detection"

Technical Report MSR-TR-2010-66, June 2010

https://publications.waset.org/15131/video-based-face-recognition-a-survey

Huafeng Wang, Yunhong Wang, and Yuan Cao

Video-based Face Recognition: A Survey

International Journal of Computer and Information Engineering, Vol. 3, No. 12, 2009

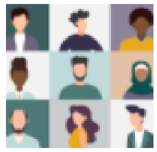
https://publications.waset.org/15131/video-based-face-recognition-a-survey

EXHIBIT C

ad exchanger

MUST READ


PUBLISHERS
Late Ad Payments Creep Back Up, But It's Not The 2020 Crisis All Over Again


DIGITAL TV AND VIDEO
Why The TV Industry Says Panels Are "In" Again

ON TV AND VIDEO

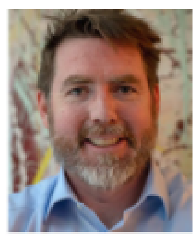
TVision Insights: 'Ratings Only Tell Part Of The Story'


By Allison Schiff

Monday, March 16th, 2020 – 12:05 am

Share:

If someone goes to the bathroom while a beautifully shot commercial plays full-screen on their TV, was it actually viewable?


Luke McGuinness, President & COO

TVISION INSIGHTS

Not so much, said Luke McGuinness, president and COO of TVision Insights, a TV analytics company that helps brands measure whether people are actually paying attention to their ads.

TVision, founded in 2014, started out by measuring attention on linear TV and moved into OTT more recently. It gathers its insights through a panel of 5,000 US homes, which represents roughly 14,000 people.

"Measuring viewability in the digital context is about whether an ad shows up in the field of view so that someone has the opportunity to see it," McGuinness said. "By and large, TV ads are showing up on the screen – but if no one is in the room to see them, they can't have an impact."

TVision clients include AB InBev, MARS, Microsoft, PepsiCo, Duracell, Google and Nestlé.

AdExchanger spoke with McGuinness.

POPULAR TODAY

PRIVACY
Oracle's Third-Party Tracking Is The Target Of A Privacy Lawsuit

ONLINE ADVERTISING
MediaMath Backs Unified ID 2.0

PODCAST: ADEXCHANGER TALKS
Fighting FOFO In The Programmatic Supply Chain

AGENCIES
The Top 10 Programmatic Agencies

COMMERCE

The CPG And Retail Relationship Status: It's Complicated

AdExchanger: TVision in a nutshell is … finish that sentence.

LUKE MCGUINNESS: We're a data and analytics company that measures how people actually watch TV. If you're an advertiser spending $100 million a year on TV advertising and people are only in the room, say, 30% of the time and only a fraction of those people are actually paying attention – well, there's a tremendous opportunity there for advertisers to better optimize and allocate their TV dollars.

How are you like Nielsen, and how are you not like Nielsen?

We are like Nielsen in that we use a panel methodology, which is 100% opt-in, but we are unlike Nielsen in that we are creating a very different metric, although our data sets are actually complementary to ratings. But ratings only tell part of the story.

We found that ratings do not correlate at all to whether people are paying attention. We've seen highly rated shows with a fairly low attention rate, and plenty of shows with a relatively small but very engaged audience and low ratings. If you're trying to optimize your advertising schedule on TV, you need to understand both. It's not just about the size of the audience but whether they are present and engaging with ads when they air.

How does your technology work?

Our panelists put our device in their homes next to their TV. It's about the size of an Apple TV and it does three key things. First, we use ACR [automatic content recognition] to determine what someone is watching on the TV. Is it "The Voice," "Stranger Things," a specific commercial?

Then we detect how the content got to the screen, whether the person is watching through live cable, the Hulu app, a Roku device, the NBCU app on Chromecast, whatever it happens to be. And then, third and most critically, our device has a camera. The technology only processes images – no video – to determine if there is anyone in the room and, if so, who. It recognizes the specific person, associates their demographics and can tell if they're paying attention to the TV or not.

How do you define attention?

Attention is our metric for engagement. We look to see whether the TV is on, which is akin to every other TV data provider out there. But then we also look at viewability, which means that the TV is on and there is at least one person in the room. Attention goes one step further: The TV is on, someone is in the room and they're actually looking at the TV. We collect this second by second but we use the MRC standard for digital video, which is two continuous seconds.

Is TVision accredited by the Media Rating Council?

We may at some point pursue accreditation, although we're not currently in the process. But we did partner with a company called Neutronian, which recently launched as a neutral auditor for data quality. They conducted a thorough audit of what we do and certified our practices as sound.

Your device relies on ACR technology. How do you deal with privacy concerns?

Our panel is 100% opt-in. Our panelists are signing nine-page contracts in order to opt in and we're very forthright in explaining how the technology works when we recruit them. We also architected our technology so that all of the data is processed on the device in the panelist's home. We're not pulling any audio or video into the cloud. Only summarized data comes back to us from the device.

How can advertisers use your data?

One way is as a complement to their existing ratings and CPM data as they head into upfront planning to optimize their TV investment for people who actually see and engage with their ads.

Some use our data to see how well attention correlates with the way they measure the outcomes of their advertising. For example, we have clients that look at brand lift on a weekly basis to measure brand health and awareness, but they never had a good way of understanding what is driving increases or decreases in awareness. Turns out when people pay attention to ads, brand lift tends to go up, and if they don't, it tends to go down. Seems logical, but they had no way to measure it.

The next logical step is to optimize for attention, for whichever networks, dayparts or specific shows their audience tends to be more engaged with. It's about allocating dollars and negotiating more effectively.

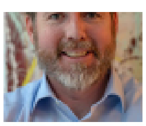
Will we have a truly data-driven upfront season this year?

Over the last couple of years, there has been the start of an evolution toward advertisers using different data sets to measure the effectiveness of their TV buys, including the amount of attention people pay to their ads. Changes won't happen overnight, but there has been an acceleration in the pace of change, which has to do, at least in part, with consumer adoption of streaming.

We're seeing more flexibility and willingness on both the buy side and the sell side to work with each other and with new data sets.

This interview has been edited and condensed.

RELATED STORIES


Luke McGuinness, President & COO

FEATURED

Nielsen's Shaky MRC Accreditation Could Accelerate Use Of Alt Currencies

Talk about it: No comments Add a comment

EXHIBIT D

T»VISION

AdAge - TVision is the go-to-choice for several Nielsen rivals

August 30 | Blog | Insights | News

*"**TVision**, a firm that originally built a panel to measure people's attention to ads, **has become the go-to choice for several Nielsen rivals** -- including VideoAmp, iSpot, Xandr and 605 -- who need person-level data to calibrate the household data they track using millions of smart TVs, set-top boxes and other devices."*

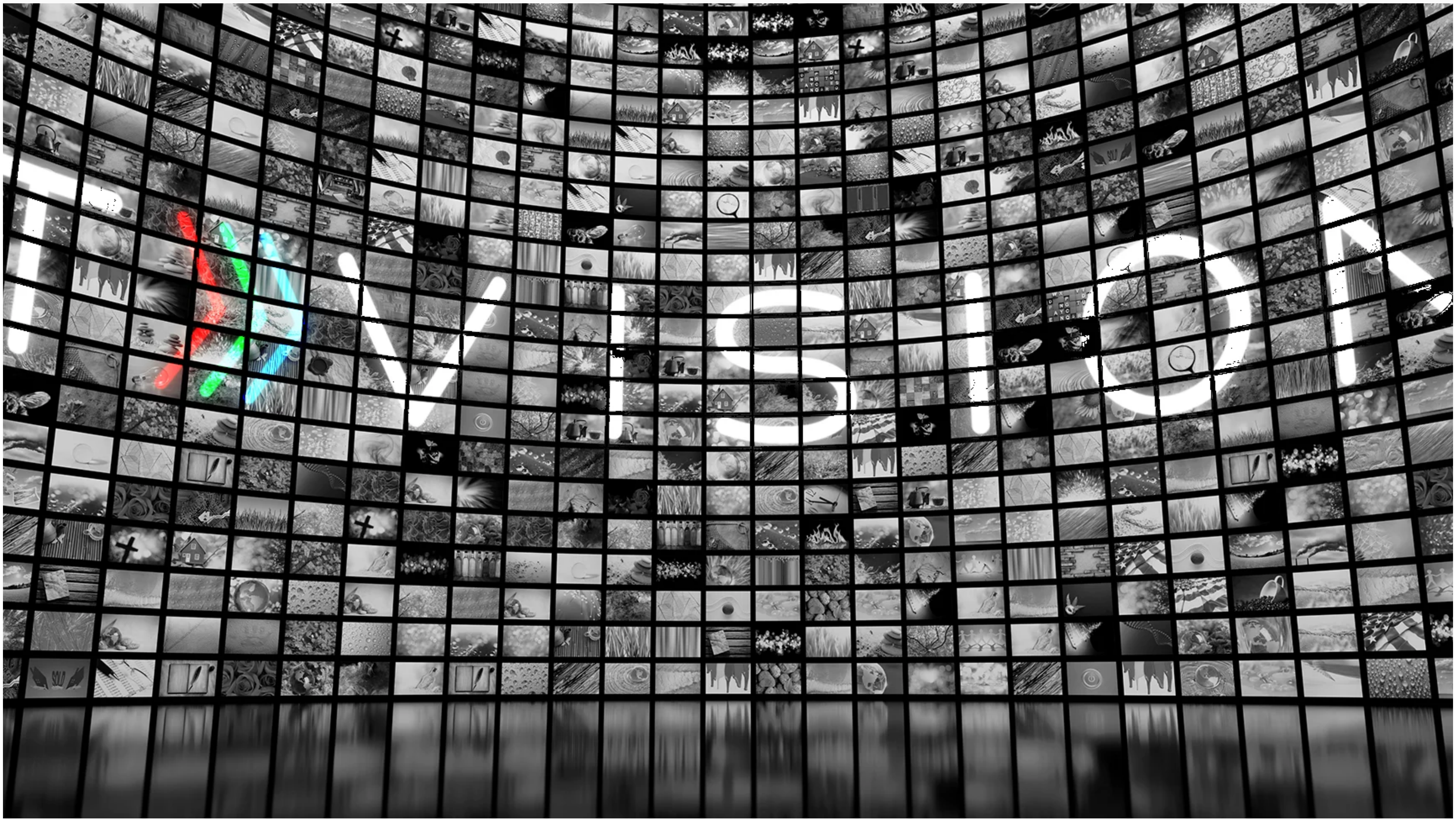
AdAge Editor Jack Neff breaks down Nielsen's decision to pause accreditation with the Media Ratings Council in his feature *"Nielsen Audience Measurement Hiatus Tests Media Ratings Council Relevance."* He examines the potential impact of that decision on the MRC, the industry, and Nielsen itself.

Neff points out that TVision is emerging as the critical partner of Nielsen competitors for person-level data. As big data rivals look to replace or augment Nielsen data as the currency standard for networks and advertisers, they are validating their census data with TVision's person-level panel data.

Neff writes, "Networks have been asking Nielsen competitors to provide person-level data, and licensing TVision's panel data could help competitors solve these problems."

Find more information about TVision's Advanced Audience Projections, person-level data solution here.

More resources from TVision



BLOG INSIGHTS

Catch Up With TVision In-Person this Fall

We are excited to get together in-person with industry leaders to discuss the future of TV and CTV and advertising attention. Catch us at TV of Tomorrow, Advertising Week, Brand Innovators, ANA Masters of Marketing, and ARF's OTT 2022.

READ MORE »

BLOG INSIGHTS

Pre Mid-Terms: Political Ads Capture Attention

As the mid-term elections approach, candidates and Super Pacs will spend an estimated $2.1 billion on linear TV and approximately $300 million on CTV advertising. Here we take a look at how effective political TV and CTV advertising is at capturing attention of the Democrat and Republican base, and engaging those crucial independent voters.

READ MORE »

INSIGHTS

The Top Ads for Attention June - August 2022

Disney+, Geico, Apple, and YouTube TV broke through to capture TV audiences' attention best between June and August, 2022. See how the ads perform with key demos and identify the creative elements that work best with this second-by-second ad analysis.

READ MORE »

Get in Touch

- [x] YES, SEND ME TVISION NEWS AND CONTENT.

- [ ] BRAND & AGENCY SOLUTIONS
- [ ] NETWORK & STREAMING SERVICE SOLUTIONS
- [ ] MEASUREMENT & DATA SOLUTIONS
- [ ] TVISION PANEL SUPPORT
- [ ] MEDIA INQUIRIES
- [ ] JOBS AT TVISION

☐ OTHER

Stay Up to Date

The TV landscape is changing quickly. Let us help keep you in the know.

FULL NAME:

EMAIL ADDRESS:*

SUBSCRIBE TO NEWSLETTER

PRIVACY POLICY

TVISION JAPAN

TVISION INDIA

EXHIBIT E

4 challenges the industry will face as it breaks away from Nielsen

campaignlive.com/article/4-challenges-industry-will-face-breaks-away-nielsen/1726140

by Alison Weissbrot September 01, 2021

The door is open for a new era of TV measurement — but can the industry thrive on multiple currencies?

When NBCUniversal last week put out an RFP for measurement partners, it was the biggest sign yet that the industry has lost faith in Nielsen.

The RFP came after Nielsen put its Media Ratings Council (MRC) accreditation on hiatus after undercounting household viewership during the pandemic. The MRC, run by George Ivie, is a trusted independent industry body formed in 1964 with a stated mission "to secure for the media industry and related users measurement services that are valid, reliable and effective."

The service is the third from Nielsen to lose MRC accreditation in the past year, including its Digital Ad Ratings (DAR) service, which it paused in October 2020, and local TV ratings, which it suspended in January.

Nielsen has placed its bets on Nielsen One, the cross-platform measurement framework it has promised to roll out in full by 2024. But stakeholders across the industry agree that 2024 is far too long to wait, given how quickly consumer viewing habits shifted during COVID, and especially now that Nielsen is no longer an accredited service.

Buyers began to move away from just using Nielsen for planning purposes years ago. VideoAmp, for example, powers Omnicom's TV planning tools, and from what I hear, is about to kick off a major test with holding companies after Labor Day to see how its currency matches up against Nielsen's. Sellers are looking for new partners as well; Comscore, for instance, is growing its remit with ViacomCBS.

Still, Nielsen gross ratings points (GRPs) remain the only independent TV currency in the market. Most buyers and sellers I've spoken with agree that NBCU's bold move is, for lack of a better term, the kick in the butt the industry needs to finally move forward.

But they're also waiting with bated breath for the massive changes, complexities and periods of confusion leading up to the new world order.

1. Juggling multiple currencies

Industry stakeholders agree that transacting on multiple currencies that correspond to advertiser business outcomes is the way forward. Jo Kinsella, president at TV measurement company TVSquared, pointed to the stock market, which trades on multiple currencies, such as futures and options, as an example. Similarly, she says, advertisers can transact on various branding and performance metrics.

Buyers are becoming more open to transacting against multiple currencies, says Michael Perlman, chief revenue officer at TVision, a TV measurement company that tracks attention. TVision recently worked with AB InBev, for example, to buy ads off of guaranteed attention metrics on A+E Network.

"There is room for multiple players," he says. "The goal of a particular campaign could dictate the right approach for deploying media."

Media buyers, however, are hesitant. Some worry that using multiple currencies will obscure the methodology to work in the networks' favor, which can choose to transact against whichever currency values their inventory the highest. Multiple currencies will also make it difficult to compare ratings across networks. And, agencies must deliver on strict pricing and savings guarantees from procurement, and multiple currencies could make it difficult to reconcile costs.

2. Getting stakeholders on board

While the industry applauds NBCU's bold stance, you'd be hard pressed to find a buyer that wants a major media seller determining TV's future currency.

Media buyers agree that any new currency adopted must receive cross-industry input and approval, as well as third-party accreditation from an independent body (most likely the MRC). Otherwise, as one buyer put it, the currency will feel "rigged." Another buyer pointed

out that it will be difficult to make the case to clients to move away from Nielsen to a new currency dictated by a media seller.

Collaboration, however, is easier said than done; getting two parties on opposite sides of a transaction to agree is tricky. Many are banking on the MRC to endorse and validate a new currency as an independent body.

But the jury is still out on whether the MRC has the appetite to create new types of accreditation for multiple currencies. As one sell-side spokesperson said, if they don't do it, "I don't know who steps in."

3. Understanding demos

Despite Nielsen's issues, it's still the largest independent panel backed by demographic household data. The Nielsen panel, though imperfect, provides a critical foundation for understanding who is watching certain types of content, beyond just what is being watched.

There are, however, rising alternatives. TVision has a 15,000-person panel that tracks not just who is in the household, but also who is watching what, using facial recognition and eye tracking technology along with automatic content recognition (ACR) data. Many Nielsen contenders are licensing TVision's dataset to underpin their systems with demographics.

As the industry slowly moves away from broad, demo-based buys in favor of outcomes, Nielsen's panel may wane in importance. As one media buyer put it, demos become arbitrary when you can measure actual sales or business goals.

4. Moving past buy-side inertia

This, in my opinion, is the most significant hurdle, and it's what has held the industry back from adopting a new currency for years. As one media buyer put it, the challenge is large and "people hope it will get solved for them."

Nielsen's historical data backs into most clients' planning processes and helps them determine volume discounts and pricing. Demand and planning haven't evolved to rightsize pricing, as clients continue to rely on old media mix modeling formulas that are no longer accurate. Even just starting to translate Nielsen currencies to a new baseline of truth would be problematic for many clients.

Another barrier is cost. Agencies spend vast amounts of money with Nielsen (one buyer estimated the costs come in just behind rent and employee salaries). If the industry were to adopt multiple currencies, costs could skyrocket, and it would be difficult for agencies, already with razor-thin margins, to offset those costs to clients.

While tough, most agree these challenges are not insurmountable. Ready or not, the next era of measurement is upon us.

Tags

Campaign Savvy Opinion

EXHIBIT F

The New York Times | https://www.nytimes.com/2017/02/25/business/media/tv-viewers-tracking-tools.html

For Marketers, TV Sets Are an Invaluable Pair of Eyes

By Sapna Maheshwari
Feb. 25, 2017

While Ellen Milz and her family were watching the Olympics last summer, their TV was watching them.

Ms. Milz, 48, who lives with her husband and three children in Chicago, had agreed to be a panelist for a company called TVision Insights, which monitored her viewing habits — and whether her eyes flicked down to her phone during the commercials, whether she was smiling or frowning — through a device on top of her TV.

“The marketing company said, ‘We’re going to ask you to put this device in your home, connect it to your TV and they’re going to watch you for the Olympics to see how you like it, what sports, your expression, who’s around,’” she said. “And I said, ‘Whatever, I have nothing to hide.’”

Ms. Milz acknowledged that she had initially found the idea odd, but that those qualms had quickly faded.

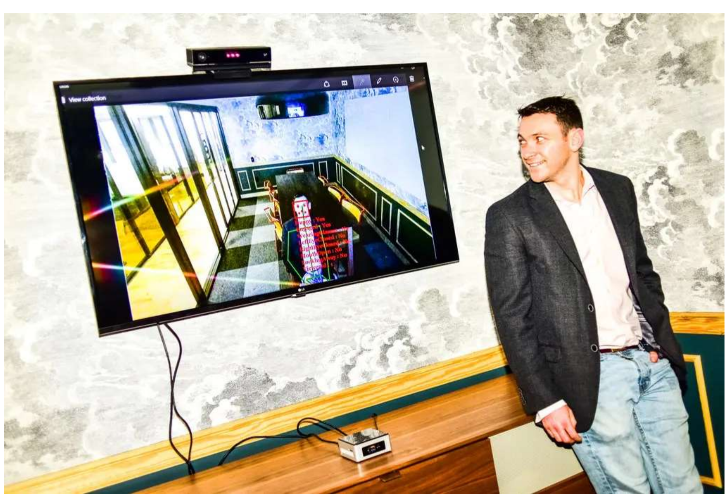
“It’s out of sight, out of mind,” she said, comparing it to the Nest security cameras in her home. She said she had initially received $60 for participating and an additional $230 after four to six months.

TVision — which has worked with the Weather Channel, NBC and the Disney ABC Television Group — is one of several companies that have entered living rooms in recent years, emerging with new, granular ways for marketers to understand how people are watching television and, in particular, commercials. The appeal of this information has soared as Americans rapidly change their viewing habits, streaming an increasing number of shows weeks or months after they first air, on devices as varied as smartphones, laptops and Roku boxes, not to mention TVs.

Through the installation of a Microsoft Kinect device, normally used for Xbox video games, on top of participants’ TVs, TVision tracks the movement of people’s eyes in relation to the television. The device’s sensors can record minute shifts for all the people in the room. The company then matches those viewing patterns to shows and commercials using technology that listens to what is being broadcast on the TV.

“The big thing for TV advertisers and the networks is: Are you actually looking at the screen or not?” said Dan Schiffman, the chief revenue officer of TVision (pronounced Tee-Vision). “What you looked at is interesting, but the fact that you looked away is arguably the most interesting.”

Mr. Schiffman founded TVision, a 30-person start-up, with a classmate from the Sloan School of Management at M.I.T.



Dan Schiffman founded TVision, a 30-person start-up, with a classmate from the Sloan School of Management at M.I.T. Dolly Faibyshev for The New York Times

Companies spend around $69 billion a year on TV ads in the United States and are keen to find out how to best distribute that money in a fractured media landscape. Nielsen and its panel of 42,500 households have long determined how money is spent on TV advertising in the United States. The higher a show's ratings, the more networks can charge for advertising.

But some industry executives have criticized Nielsen's methods as outdated. Nielsen selects homes at random to represent the nation's viewing audience, and measures who is watching what shows, mostly through meters connected to the sets, as well as diaries in select markets and digital tracking of certain ad-supported programs on tablets and phones.

The company recently delayed the rollout of a new system that will count viewing across platforms and devices. The capability to do just that is a core selling point for upstarts like TVision, which promote their ability to measure how people are binge-watching shows on, say, Netflix and Amazon.

"Nielsen will remain the currency for the time being because it is agreed upon as the thing everyone uses," said Alan Wurtzel, an adviser at NBCUniversal and its former head of research. "But as the world becomes more complex, as it is, many more additional supplemental or complementary measures will come into play."

Information gathered by companies like TVision can help advertisers steer marketing toward shows with the most engaged audiences, not just the largest ones. And for networks, it could make a show with a committed and loyal audience as valuable as one that attracts a larger but more casual set of viewers.

TVision has recruited 2,000 households, or roughly 7,500 people, in the Boston, Chicago and Dallas-Fort Worth areas. The company said the information was transmitted without storing images or video and collected anonymously.

Mr. Schiffman said the data would show, for example, "Person No. 124 in Household 6 was paying attention this second and not paying attention the next to a certain program or advertisement."

Symphony Advanced Media has built a panel of 17,500 people in the United States who have installed its Media Insiders mobile app, mostly on Android phones. In exchange for about $5 to $12 a month, the app passively tracks how people use their phones, uses the device's microphone to hear what they are watching and asks them to complete surveys. The app can tell if someone streamed a show using headphones while on a bus or saw a commercial at a sports bar.

Symphony and TVision both use technology that can recognize shows and ads through the audio or digital tags the content contains. Consumers are most likely familiar with this type of technology through the song-recognition app Shazam.

A camera on top of a television tracks the movement of people's eyes in relation to the screen. Dolly Faibyshev for The New York Times

Another ratings company, RealityMine, has assembled a panel of 5,000 people in the United States whom it said it paid less than $90 a year, who either have its app, a "home meter" plugged into their internet networks or both. In some instances, its meter may capture activity across 25 devices in a household, such as tablet, phone, Xbox, Wii, Apple TV and Google Chromecast.

The aim is "understanding what is the media day and the life of the consumer today," said Charlie Buchwalter, the chief executive of Symphony, which has worked with companies including NBCUniversal.

Mr. Schiffman said that while some people were wary of TVision's technology, they were often placated after learning that it was not storing images or videos.

During the Olympics, Ms. Milz wore a Fitbit so that NBC could see how her heart rate changed while she watched certain events.

"We're just trying to understand where people really are and what they're doing, what they're watching, how are they interacting, and ideally after that, how is that changing their behavior or affecting their behavior," said Jonathan Steuer, the chief research officer of Omnicom Media Group, which oversees media buying for advertisers.

Still, privacy can be a concern. This month, Vizio, one of the biggest makers of internet-connected televisions, said it would pay $2.2 million to settle charges that it had been collecting and selling viewing data from millions of smart TVs without the knowledge or consent of the sets' owners.

By measuring the level of attention a person is paying to a given show, TVision believes it can help bolster niche programs and smaller networks. For example, Mr. Schiffman said the company had found that the series "Lucifer," on Fox, commanded better attention metrics from viewers than "The Big Bang Theory," on CBS, even though "Big Bang" is one of the TV shows rated highest by Nielsen.

"People don't just tune into 'Lucifer.' They DVR it and watch it when they come home," he said, adding that viewers tend to be focused on the show and stay in the room when it is on.

The Weather Channel used TVision's data in the fall to give it an edge over the news and lifestyle shows it is normally compared against by advertisers. It showed its audience for weather news as "lean forward, lean-in viewers," said Indira Venkat, who oversees research at the channel.

"What is oftentimes missing is the quality of the audience," Ms. Venkat said about Nielsen viewing data. "Yes, you're getting audience, but what are they doing in today's era of multitasking?"

EXHIBIT G

For TV's Ad Future, All Eyes Are On Attention Metrics

By **Adam Jacobson** - September 26, 2017

Dan Schiffman, Chief Revenue Officer of TVision

RBR+TVBR INFOCUS

Picture this: After an exhausting day, you're plopped yourself down on the couch and finally have time to see the series premiere of FOX's *The Orville*. About 15 minutes in, you're fast asleep, snoozing away while ignoring what will likely be a huge turkey from *Family Guy* creator and Frank Sinatra Sr. superfan Seth MacFarlane.

Nielsen may have a number to give FOX, because you "watched" it. New technology **TVision Insights** proves otherwise. What does this mean for the future of TV, and for "attention metrics"? TVision Insights Chief Revenue Officer **Dan Schiffman** spoke with RBR+TVBR from his Union Square offices to explain exactly how CMOs in charge of media budgets can adapt and embrace his company's technology. At the same time, TV networks and their affiliate stations have earned a wake-up call with respect to who is actually paying attention to their programming — and their commercials.

Adam Jacobson

Adam R Jacobson is a veteran radio industry journalist and advertising industry analyst with general, multicultural and Hispanic market expertise. From 1996 to 2006 he served as an editor at Radio & Records.

EXHIBIT H

TVISION

The Future of Media Measurement: The Role of Panels in Big Data

By Yan Liu, CEO | June 16 | Blog | Insights

As audiences increasingly migrate to CTV, the old norms of TV measurement no longer apply. A new approach is required - one that is best served through a combination of big data and calibration panels. While smaller in size than big data, the importance of calibration panels' role in ensuring the accuracy of future measurement solutions should not be underestimated. So what role will the panel play? This is a question we are asked often, afterall, we are one of the industry's only person-level panels. I was happy to get to the heart of this question at the Summer TV Data and Measurement Summit sponsored by NextTV where I discussed the future of the panel with industry gurus Nielsen's Mainak Mazumdar, Beth Rockwood of WarnerMedia, Peter Sedlarcik from Havas Media Group, and Howard Shimmel of Janus (former CRO of Turner). We all agreed panels are pivotal to the future of measurement.

As the WFA outlines in their Cross-Media Measurement Framework, effectively measuring media across digital, TV and CTV is a complex problem that will take multiple data sets to solve. Both census-level big data and panels will be

part of the solution working together to provide the most comprehensive measurement of audience behavior and engagement with all types of media, whether digital or TV.

In this new model, big data is the foundation and panels are just one ingredient. The simple truth is that big data works better if it is calibrated by panels. Calibration panels can, and should be used to fine tune census-level data to remove biases, fill in gaps, and ensure accuracy by acting as a common source of truth.

What Makes a Calibration Panel Effective for Unified Media Measurement

At TVision, we've already begun supporting several big data providers who are looking to develop next-gen measurement solutions. We've found that in order for panels to work effectively as calibration data sets, as our does, they must meet these criteria:

- **Person-level data** - At TVision, we have a saying, "Households don't buy products, people do." Any effective measurement solution must accurately report the marketer's ability to reach specific decision-makers, and not a generic household. Person-level is a critical panel requirement because, unfortunately, a significant percentage of big data sets still rely on household-level information. Person-level data from panels enable big data sets to normalize their reported viewer behavior based on who is actually watching. For example, household-level data might show ESPN to have an equal number of male and female viewers - assuming an equal number of males and females live in the household. Person-level insights from a panel can help calibrate the big data set to learn what percentage of ESPN viewers are typically male or female.
- **Second-by-second analysis** - With video commercial air times commonly less than 30 seconds, and increasingly just six seconds, the industry needs insight into second-by-second viewer behavior in order to effectively measure video ads. If data is only available on a minute-by-minute basis, measurement companies may assume a viewer has seen an ad when in fact they have tuned away. Second-by-second data is critical for accurately reporting metrics for ad delivery and engagement.
- **Single-source across devices -** Single-source panels measure the behavior of the same individuals across devices, regardless of if they are viewing linear or CTV. They enable apples-to-apples analysis and provide a comprehensive view of how people are consuming media. The WFA framework sums it up nicely: a single-source panel "acts as the arbiter of truth, providing benchmarks for the use and overlap of media consumption across channels and screens."
- **Passive measurement** - Measuring passive panelist behavior enables continuous, accurate understanding of viewer behavior. There is little opportunity for human error or biases to impact the data. For instance, passive measurement helps provide correct co-viewing metrics since it doesn't rely on the panelists to self-identify that they are in the room. Passive measurement can also be trusted to provide more accurate person-level identification of guests, children, and other non-heads of household.
- **Transparency** - Panel data shouldn't be a black box. If assumptions or adjustments have been made within the panel data in order to provide statistical significance, it needs to be clear what those are so they can be factored into the combined model. Upfront transparency is key for removing any bias that could be exaggerated as it gets scaled up within the larger data set.
- **Flexibility** - The future of measurement is complex. We need the entire industry to work together to solve our common problems. While some measurement companies may be looking for an out-of-the-box solution, innovation will be more likely if companies can manipulate the data on their own. For providers who want to focus on the unique value their solution provides, a one-size-fits-all panel data set will be too limiting. If panels provide access to sessionized data, the data scientists at these measurement companies can more efficiently work with the panel data to get what they need for their own models.

A Future with Multiple Currencies for Media Measurement

This is an exciting time for our industry. We all know we're in desperate need of new currencies, ones that place value on the effectiveness of ads. While there is certainly a continued place for media ratings, I propose that we work toward a future that supports multiple currencies. After all, different marketers have different goals for their campaigns. Brand marketers care about attention, whereas direct marketers are more focused on immediate conversions. To align success, these varying goals need to be supported with currencies that are built into the campaign from media planning to measurement.

As TV embraces big data, every company should have as much runway as possible to contribute to our new path forward. Measurement providers shouldn't be boxed into one panel provider with one vision of measurement - there is room for multiple panels as we redefine how we measure and value media. Let's not forget, a multi-panel ecosystem is good not only for innovation, but also as a check and balance to ensure the wrong trends don't go unnoticed, or worse, expanded upon as big data is modeled up.

There is plenty of room for innovation and improvement. If our industry can unlock the power of big data, we can solve numerous problems. We don't need to have all our eggs in one basket, and frankly, the cost of panel data shouldn't eat up all the research budgets that could otherwise drive innovation. Our industry's best future will unfold when measurement providers have flexibility in their approach as they work to harness the power of big data and the accuracy of panels together.

More resources from TVision

BLOG INSIGHTS

Catch Up With TVision In-Person this Fall

We are excited to get together in-person with industry leaders to discuss the future of TV and CTV and advertising attention. Catch us at TV of Tomorrow, Advertising Week, Brand Innovators, ANA Masters of Marketing, and ARF's OTT 2022.

READ MORE »

BLOG INSIGHTS

Pre Mid-Terms: Political Ads Capture Attention

As the mid-term elections approach, candidates and Super Pacs will spend an estimated $2.1 billion on linear TV and approximately $300 million on CTV advertising. Here we take a look at how effective political TV and CTV advertising is at capturing attention of the Democrat and Republican base, and engaging those crucial independent voters.

READ MORE »

INSIGHTS

The Top Ads for Attention June - August 2022

Disney+, Geico, Apple, and YouTube TV broke through to capture TV audiences' attention best between June and August, 2022. See how the ads perform with key demos and identify the creative elements that work best with this second-by-second ad analysis.

READ MORE »»

Get in Touch

- [x] YES, SEND ME TVISION NEWS AND CONTENT.

- [ ] BRAND & AGENCY SOLUTIONS
- [ ] NETWORK & STREAMING SERVICE SOLUTIONS
- [ ] MEASUREMENT & DATA SOLUTIONS
- [ ] TVISION PANEL SUPPORT
- [ ] MEDIA INQUIRIES
- [ ] JOBS AT TVISION
- [ ] OTHER

Stay Up to Date

The TV landscape is changing quickly. Let us help keep you in the know.

FULL NAME:

EMAIL ADDRESS:*

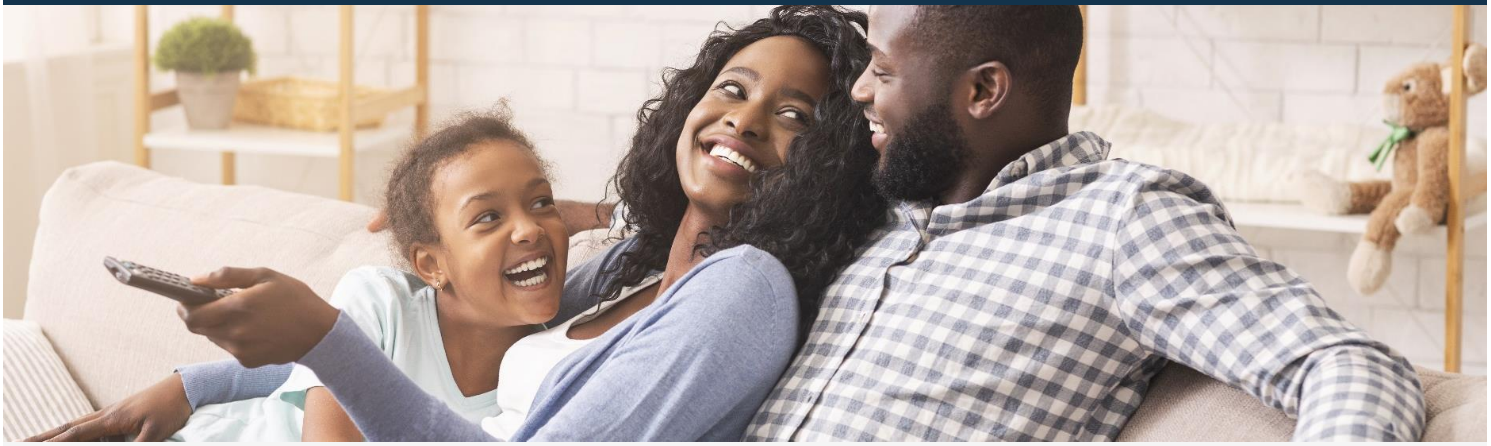
SUBSCRIBE TO NEWSLETTER

PRIVACY POLICY
TVISION JAPAN
TVISION INDIA

EXHIBIT I

TVISION

Advanced Audience Projections



TVision Advanced Audience Projections Powers Person-Level Ad Measurement

Advertisers are demanding more clarity in the TV measurement data. Go beyond household data from set-top boxes and Smart TVs to deliver person-level insights into how people really watch TV across linear and streaming. Meet the growing industry call for cross-platform measurement with person-level TV viewership and Attention data that matches digital-style clarity.

Tvision AAP is a flexible solution that allows measurement partners to maximize the benefits of the combination of their own data and TVision's unique person-level insights. Join the leading TV and advertising platforms already using Tvision's data to deliver more powerful insights:

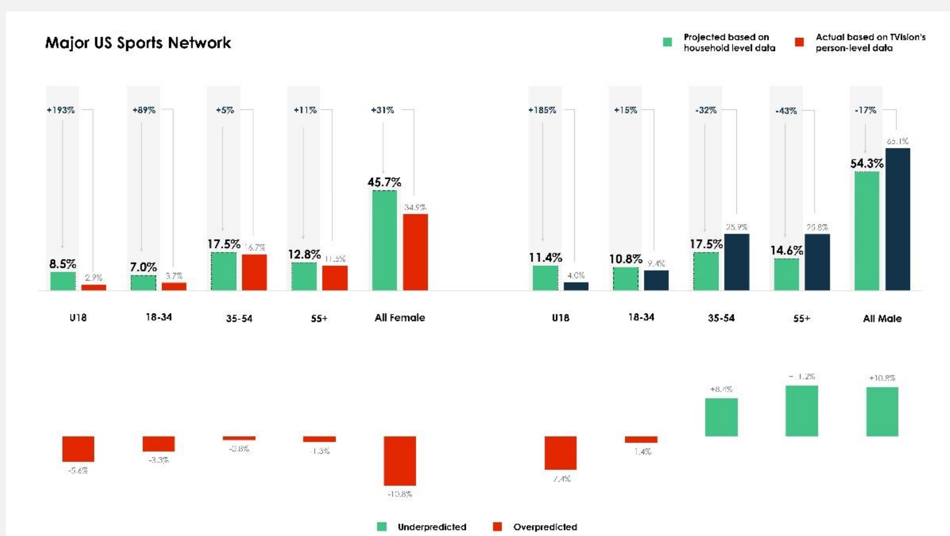
- TV Attribution - Tie person-level ad exposure data to outcome data with significantly greater accuracy.
- Co-Viewing - Quantify the value of co-viewing in the room when the ad and content aired.
- Reach and Frequency - Understand unduplicated reach and frequency as well as incremental reach for specific viewers.
- Cross-Platform Measurement - Meet MRC and WFA standards for comparative measurement across TV and digital.

Households do not watch TV.
People watch TV.

Tvision's data gives the industry the clearest picture of who is actually watching TV.

The difference between measuring TV at the household-level and the person-level is dramatic and it can cost advertisers millions of dollars in wasted spending.

By knowing exactly who views the content and ad, the industry can more effectively value content.



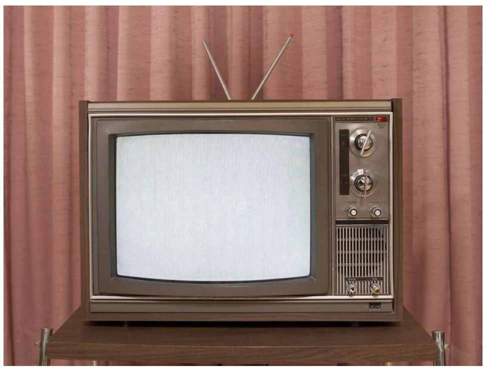
ORACLE xandr videoamp iSpot.tv

EXHIBIT J

TVision raises $6.8M to take on Nielsen with thermal eye and emotion tracking tech

Ingrid Lunden @ingridlunden / 9:36 AM EDT • October 26, 2016

Comment



Another startup out of MIT built on computer vision and focused on eye-tracking technology has raised funding to build out its business: Boston-based TVision Insights, which tracks who is watching what on TV and how they are reacting to it, and then works with advertisers and broadcasters to provide them with that data to have better insights into their programming, has raised $6.8 million.

The funding — which brings the total raised by TVision to $9.65 million — comes from Accomplice (formerly known as Atlas Venture), along with Golden Venture Partners, Jump Capital, and ITOCHU Technology Ventures (which has backed the likes of Box but also Fab, among many more startups).

There are a lot of startups right now gaining attention for the way that they are using advances in computer vision and machine learning to track what your eyes are doing. Just yesterday, it was announced that Google acquired Eyefluence, most likely to boost its efforts in emerging areas like virtual reality. Another startup that came out of MIT, Affectiva, started out focusing on emotional responsiveness to online videos and has more recently made some interesting inroads into robotics and automotive applications.

TVision is doing something a little different from these and has been built specifically to address the gap in how TV viewing is measured, CRO Dan Schiffman — who co-founded the company with CEO Yan Liu, Pongpun Pong Laosettanun, Alex Amis and Raymond Fu — explained to me.

The problem that TVision is solving is a well-known one in the TV world. There are a number of companies like Nielsen that already measure TV viewing, but many of them simply monitor when the TV is on, relying on the users themselves to indicate who is watching and when, and who is actually watching the TV rather than sitting on the sofa and playing on their phones instead. Variables like these can result in data that is not completely accurate.

And at a time when digital platforms are all about providing viewing data, and many users are already migrating away from tradition TV viewing, that reporting shortfall could eventually lead to

advertising declines in a medium that has dominated advertising for decades but is facing a lot of competition from newer platforms like social media, mobile and streamed video.

"TVision provides an important solution for next-level analysis to an industry that is desperate for new and better ways to measure audience attention," said Ryan Moore, partner and founder at Accomplice, in a statement. "Yan and his team offer the TV industry and advertisers a solution to make high-value programming and advertising decisions based on data they simply did not have before."

The company starts with a small device that sits on top of and works with your ordinary television. It does not read the world as we see it with an optical camera alone; it uses lasers and thermal infra-red so that it can pick up more data even when lighting conditions are low (as they often are when you are watching TV). Its sensors and algorithms are capable of identifying not just who is watching in a family group, but also who is just sitting in the room but not watching TV (instead playing on, say, a mobile phone), and what viewers' reactions are to the show that is on at the time.

As Schiffman describes it, "we than translate all that data into ones and zeros, and figure out how to make sense of it."

TVision's devices are currently installed in some 7,000 homes in the U.S. and Japan as part of an opt-in, Nielsen-style panel, Schiffman said. The idea is to use some of the funding for business development to bring that number up to 15,000.

The startup already provides data to three of the largest broadcasters in the U.S., as well as many major advertsiers — although these are under NDA and so the names cannot be disclosed, Schiffman said.

Some of the funding will also go towards hiring more talent to expand beyond its current 19 employees, as well as for R&D. Schiffman told me that TVision already has applications in for two utility patents, one for its computer vision algorithm and another around its analytics.

Join TechCrunch+
Login
Search
TC Disrupt 2022
Startups
TechCrunch+
Podcasts
Newsletters
Startup Battlefield
Advertise
Events
More

More TechCrunch

Deepfakes for all: Uncensored AI art model prompts ethics questions

Google research AI image noise reduction is out of this world

CEO and COO of Pornhub parent company MindGeek abruptly resign

When did the startup boom peak?

- [ ] TechCrunch+ Announcements
- [ ] TechCrunch+ Events
- [ ] TechCrunch+ Roundup

See all newsletters

Email *

Subscribe

http://tcrn.ch/2f7bQJz Copy

Tags

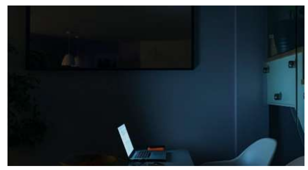
accomplice ad tech Advertising Tech Computer Vision emotional analysis eye tracking

TVision Insights

Join TechCrunch+
Login
Search
TC Disrupt 2022
Startups
TechCrunch+
Podcasts
Newsletters
Startup Battlefield
Advertise
Events
More

An experimental new attack can steal data from air-gapped computers using a phone's gyroscope

Carly Page
4:00 PM EDT • August 24, 2022



TechCrunch Disrupt 2022

Check out the official Disrupt 2022 roundtables

Lauren Simonds
4:00 PM EDT • August 24, 2022

SiriusXM digs into subscription-based podcast market with Stitcher Premium bundle

Andrew Mendez
3:31 PM EDT • August 24, 2022

TC+ Fundraising

Ron DeSantis' proposed ban on ESG investments would be another blow for diverse fund managers

Dominic-Madori Davis
3:00 PM EDT • August 24, 2022

Shuffles, Pinterest's invite-only collage-making app, is blowing up on TikTok — here's how to get in

Sarah Perez
2:28 PM EDT • August 24, 2022

Amazon's adding voice-powered gaming to big name games with Alexa Game Control

Lauren Forristal
2:27 PM EDT • August 24, 2022

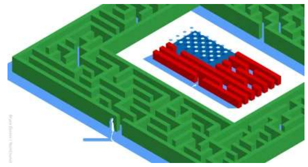
India lifts ban on American Express

Manish Singh
2:08 PM EDT • August 24, 2022

Join TechCrunch+
Login
Search
TC Disrupt 2022
Startups
TechCrunch+
Podcasts
Newsletters
Startup Battlefield
Advertise
Events
More

Instacart now lets you order same-day delivery for large items, including furniture and electronics
Aisha Malik
2:02 PM EDT • August 24, 2022

TC+ Work
Dear Sophie: What should we know about the H-1B lottery before we hire STEM OPTs?
Sophie Alcorn
2:00 PM EDT • August 24, 2022



See the Series B pitch deck that lead to Benchling's $6B valuation on today's TechCrunch Live
Matt Burns
1:56 PM EDT • August 24, 2022

Twitter faces privacy scrutiny from EU watchdogs after Mudge report
Natasha Lomas
1:22 PM EDT • August 24, 2022

TechCrunch+ Work
ARR per employee is the North Star efficiency metric you've been looking for
Neeraj Agrawal, Brandon Gleklen, Jack Mattei
1:00 PM EDT • August 24, 2022

Black Girls Code's developing story offers a complicated look at lots of different things
Natasha Mascarenhas, Dominic-Madori Davis, Alex Wilhelm
12:31 PM EDT • August 24, 2022

Apple will debut the iPhone 14 on September 7

Brian Heater
12:02 PM EDT • August 24, 2022

Join TechCrunch+
Login
Search
TC Disrupt 2022
Startups
TechCrunch+
Podcasts
Newsletters
Startup Battlefield
Advertise
Events
More

TC+ Market Analysis

Former US CFTC commissioner says crypto space has 'dire need for regulatory clarity'

Jacquelyn Melinek
12:00 PM EDT • August 24, 2022

Lily AI lands new capital to help retailers match customers with products

Kyle Wiggers
12:00 PM EDT • August 24, 2022

TechCrunch Disrupt 2022

Serena Williams will discuss her expanding focus on VC at TC Disrupt

Lauren Simonds
11:30 AM EDT • August 24, 2022

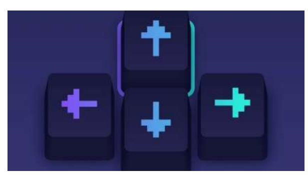
Snap agrees to $35 million settlement in Illinois privacy lawsuit

Aisha Malik
11:19 AM EDT • August 24, 2022

TC+ Work

How to incorporate web3 gaming into your Web 2.0 strategy

Tom Fleetham
11:00 AM EDT • August 24, 2022



Amazon and DirecTV partnership brings 'Thursday Night Football' to 300,000+ bars, restaurants and other venues

Lauren Forristal
10:57 AM EDT • August 24, 2022

About
TechCrunch
Staff
Contact Us
Advertise
Crunchboard Jobs

Legal
Privacy Policy
Terms of Service
TechCrunch+ Terms
Privacy Dashboard
Code of Conduct
About Our Ads

Facebook
Twitter
YouTube
Instagram
LinkedIn

Join TechCrunch+
Login

Search

TC Disrupt 2022
Startups
TechCrunch+
Podcasts
Newsletters
Startup Battlefield
Advertise
Events
More

EXHIBIT K

US 20180007431A1

(19) United States
(12) Patent Application Publication
(10) Pub. No.: US 2018/0007431 A1
(43) Pub. Date: Jan. 4, 2018

Sidhu et al.

(54) SYSTEMS AND METHODS FOR ASSESSING VIEWER ENGAGEMENT

(71) Applicant: **TVision Insights, Inc.**, Boston, MA (US)

(72) Inventors: **Inderbir Sidhu**, Lexington, MA (US); **Yanfeng Liu**, Long Island City, NY (US); **Yun Fu**, Brookline, MA (US)

(21) Appl. No.: **15/702,229**

(22) Filed: **Sep. 12, 2017**

Related U.S. Application Data

(63) Continuation of application No. PCT/US2017/012531, filed on Jan. 6, 2017.

(60) Provisional application No. 62/275,699, filed on Jan. 6, 2016.

Publication Classification

(51) **Int. Cl.**

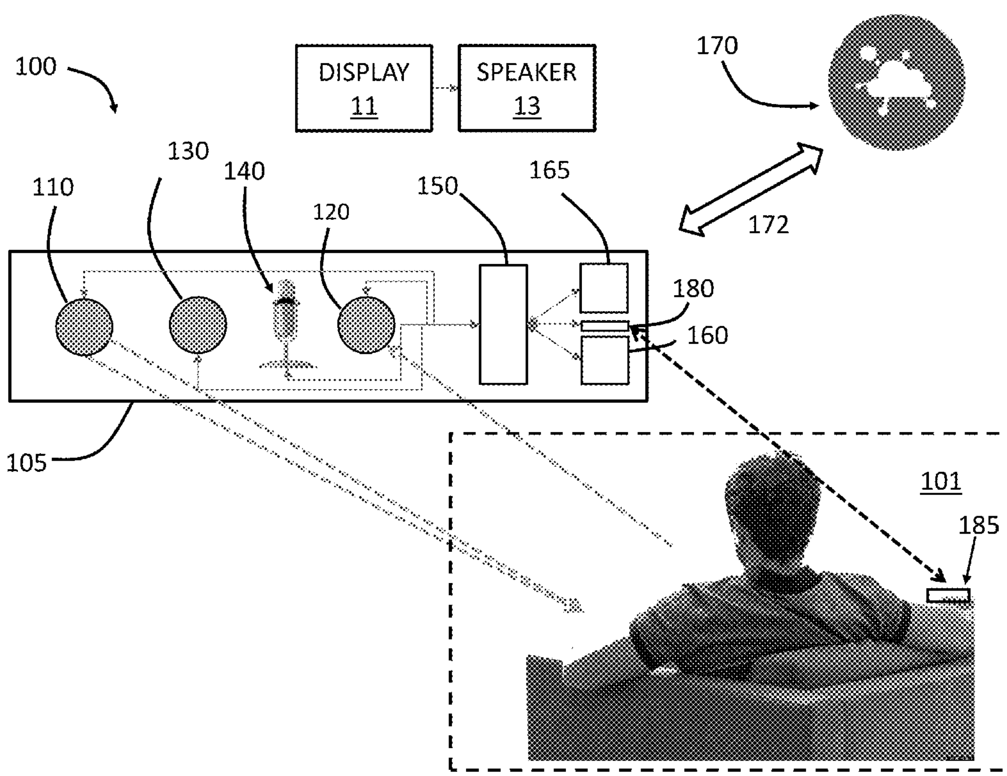
H04N 21/442	(2011.01)
H04N 21/422	(2011.01)
H04N 21/4223	(2011.01)
H04N 21/439	(2011.01)
H04N 21/437	(2011.01)
H04N 5/33	(2006.01)
G06K 9/00	(2006.01)
G06K 9/20	(2006.01)
H04N 5/247	(2006.01)

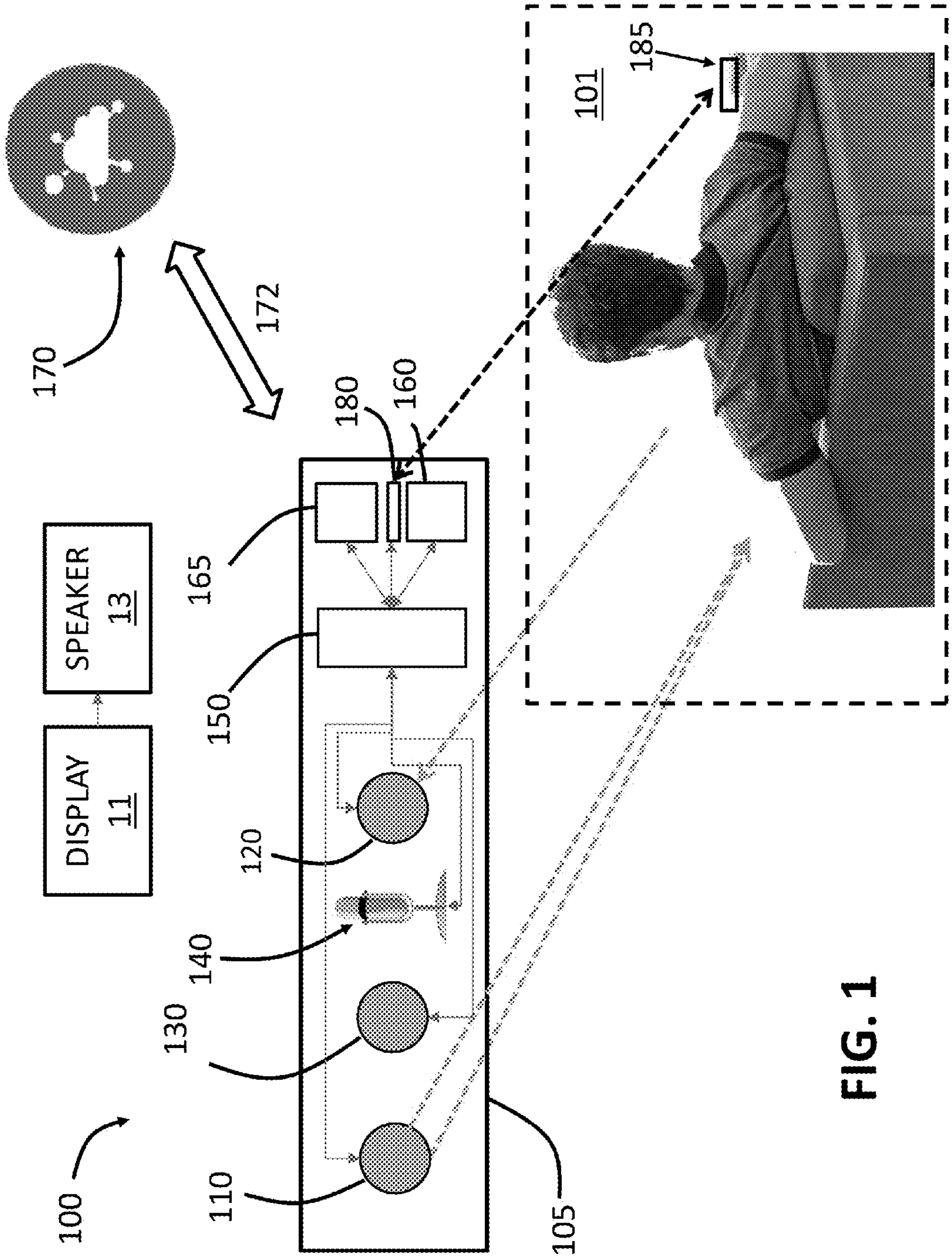
(52) **U.S. Cl.**

CPC ... ***H04N 21/44218*** (2013.01); ***H04N 21/4394*** (2013.01); ***H04N 21/42203*** (2013.01); ***H04N 21/4223*** (2013.01); ***H04N 21/437*** (2013.01); *G06K 9/00597* (2013.01); *G06K 9/00369* (2013.01); *G06K 9/00302* (2013.01); *G06K 9/00288* (2013.01); *G06K 9/00281* (2013.01); *G06K 9/2018* (2013.01); *G06K 9/00228* (2013.01); *H04N 5/247* (2013.01); *H04N 5/33* (2013.01); *G06K 2009/00322* (2013.01)

(57) **ABSTRACT**

A system for quantifying viewer engagement with a video playing on a display includes at least one camera to acquire image data of a viewing area in front of the display. A microphone acquires audio data emitted by a speaker coupled to the display. The system also includes a memory to store processor-executable instructions and a processor. Upon execution of the processor-executable instructions, the processor receives the image data and the audio data and determines an identity of the video displayed on the display based on the audio data. The processor also estimates a first number of people present in the viewing area and a second number of people engaged with the video. The processor further quantifies the viewer engagement of the video based on the first number of people and the second number of people.





100
110
130
140
120
150
165
180
160
105
170
172
DISPLAY
11
SPEAKER
13
101
185

FIG. 1

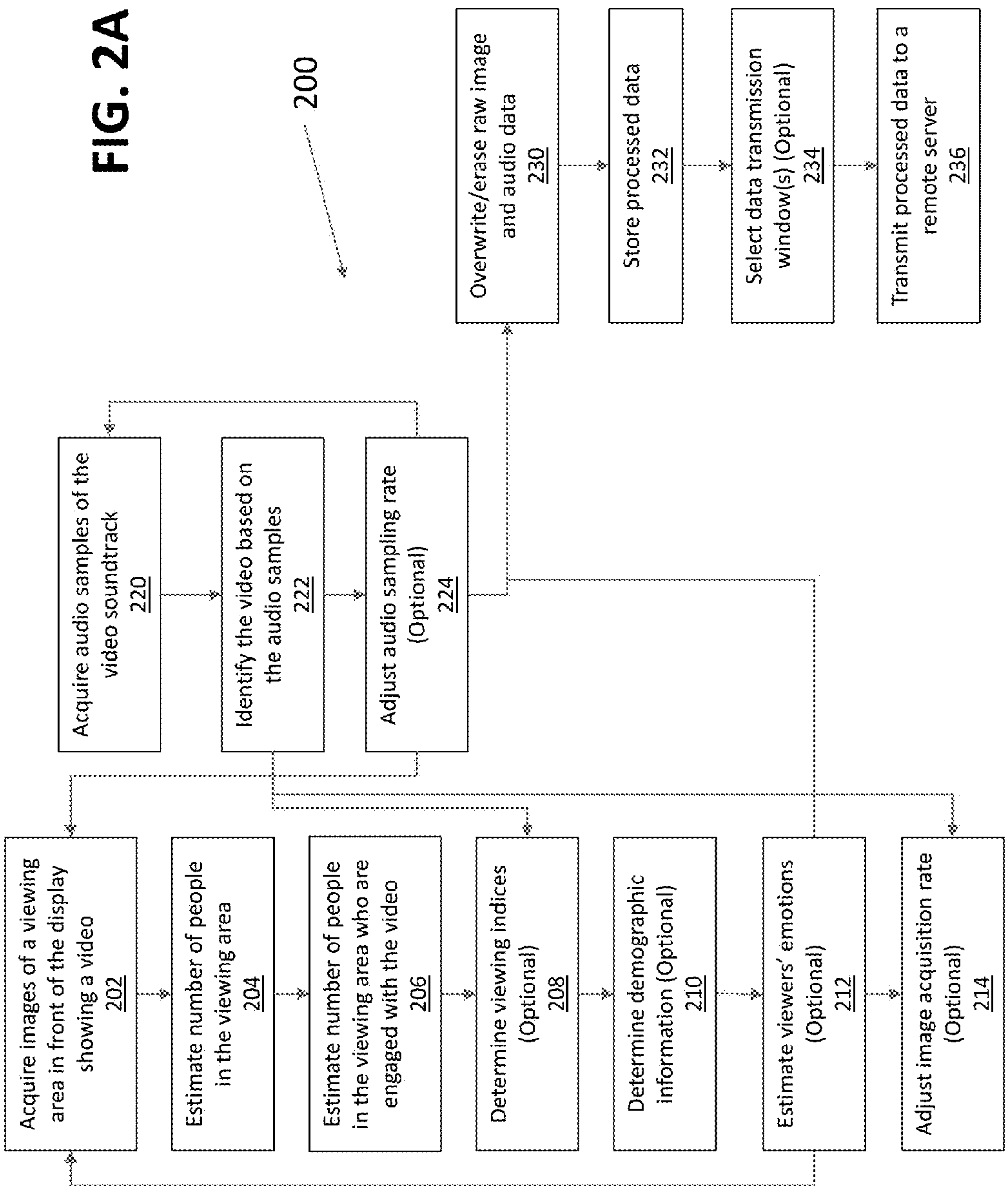

200
Acquire images of a viewing area in front of the display showing a video 202
Estimate number of people in the viewing area 204
Estimate number of people in the viewing area who are engaged with the video 206
Determine viewing indices (Optional) 208
Determine demographic information (Optional) 210
Estimate viewers' emotions (Optional) 212
Adjust image acquisition rate (Optional) 214
Acquire audio samples of the video soundtrack 220
Identify the video based on the audio samples 222
Adjust audio sampling rate (Optional) 224
Overwrite/erase raw image and audio data 230
Store processed data 232
Select data transmission window(s) (Optional) 234
Transmit processed data to a remote server 236

FIG. 2A

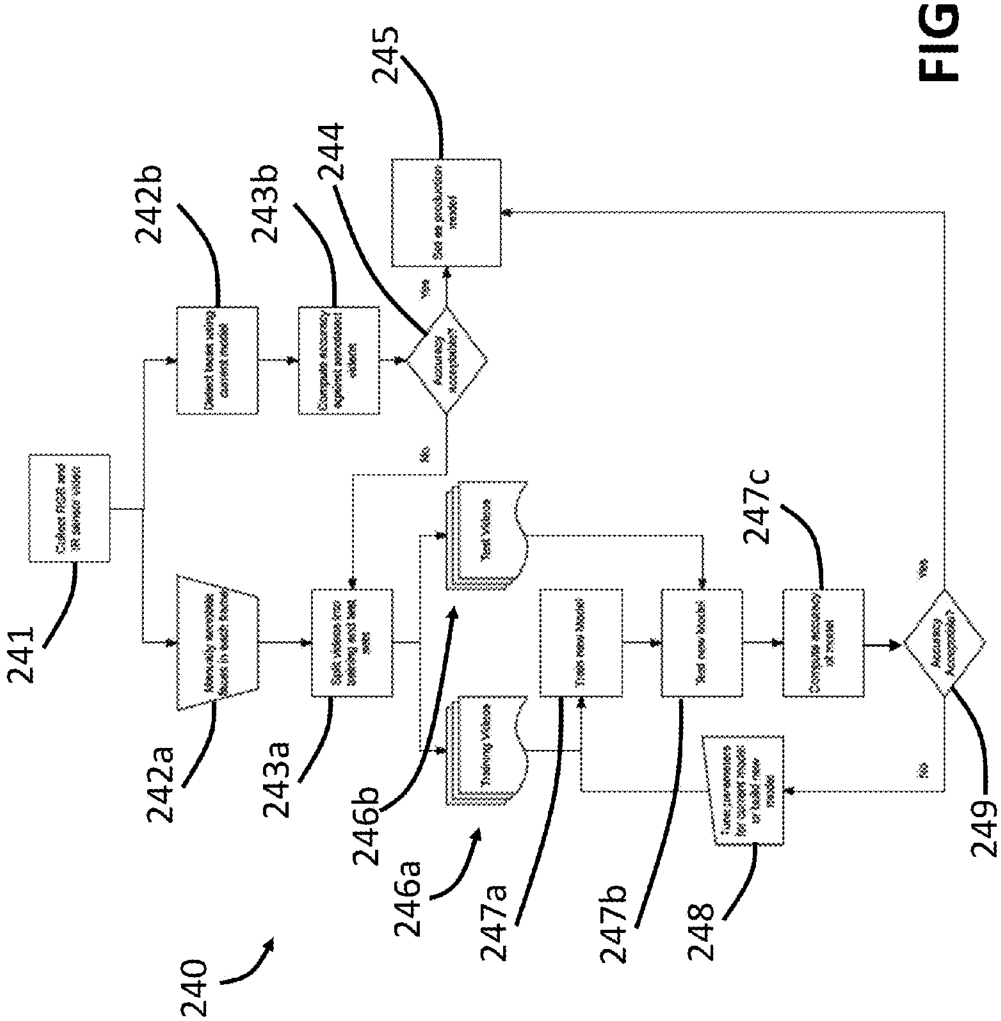

240
241
242a
242b
243a
243b
244
245
246a
246b
247a
247b
247c
248
249

FIG. 2B

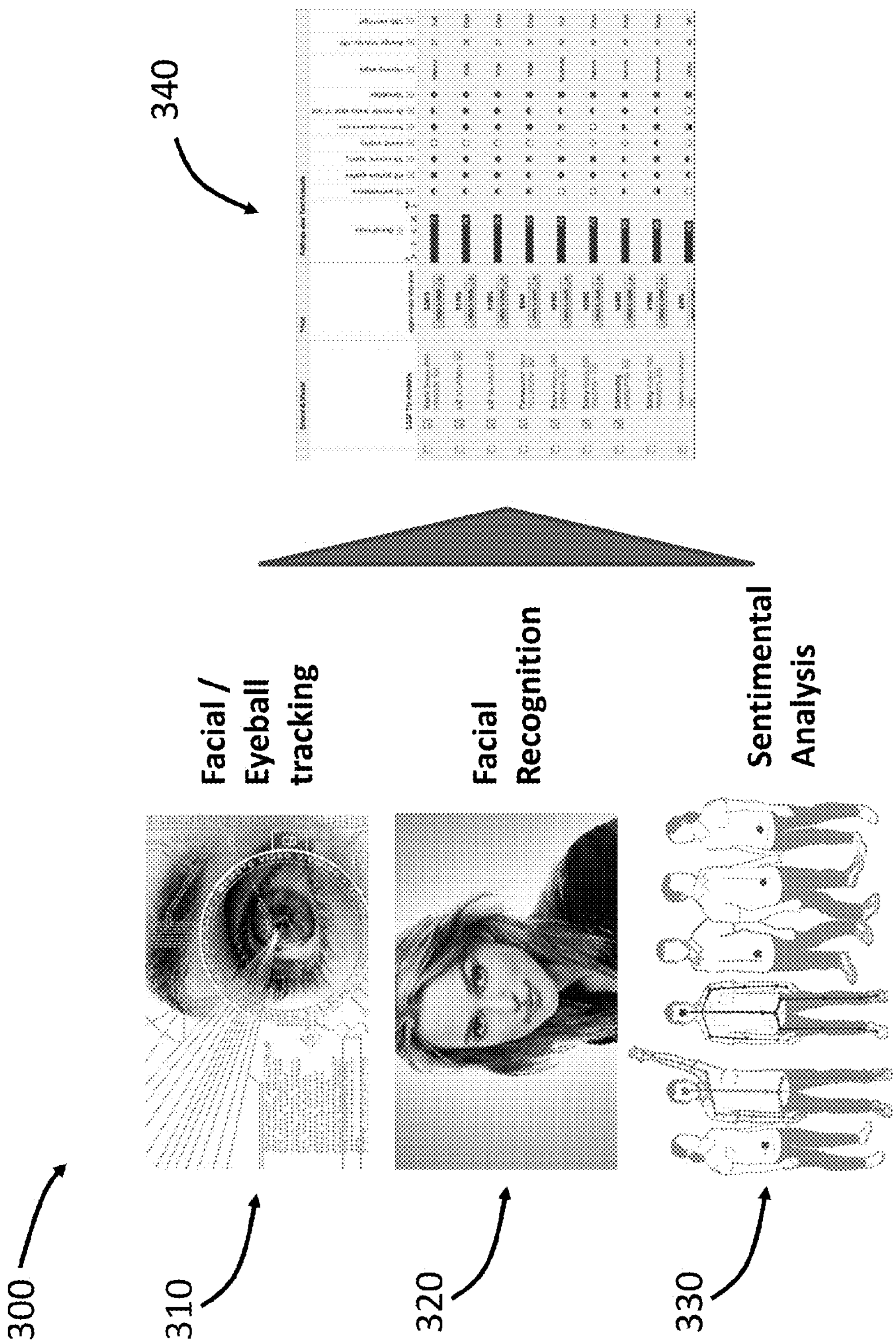

340
Facial / Eyeball tracking
Facial Recognition
Sentimental Analysis
300
310
320
330

FIG. 3A

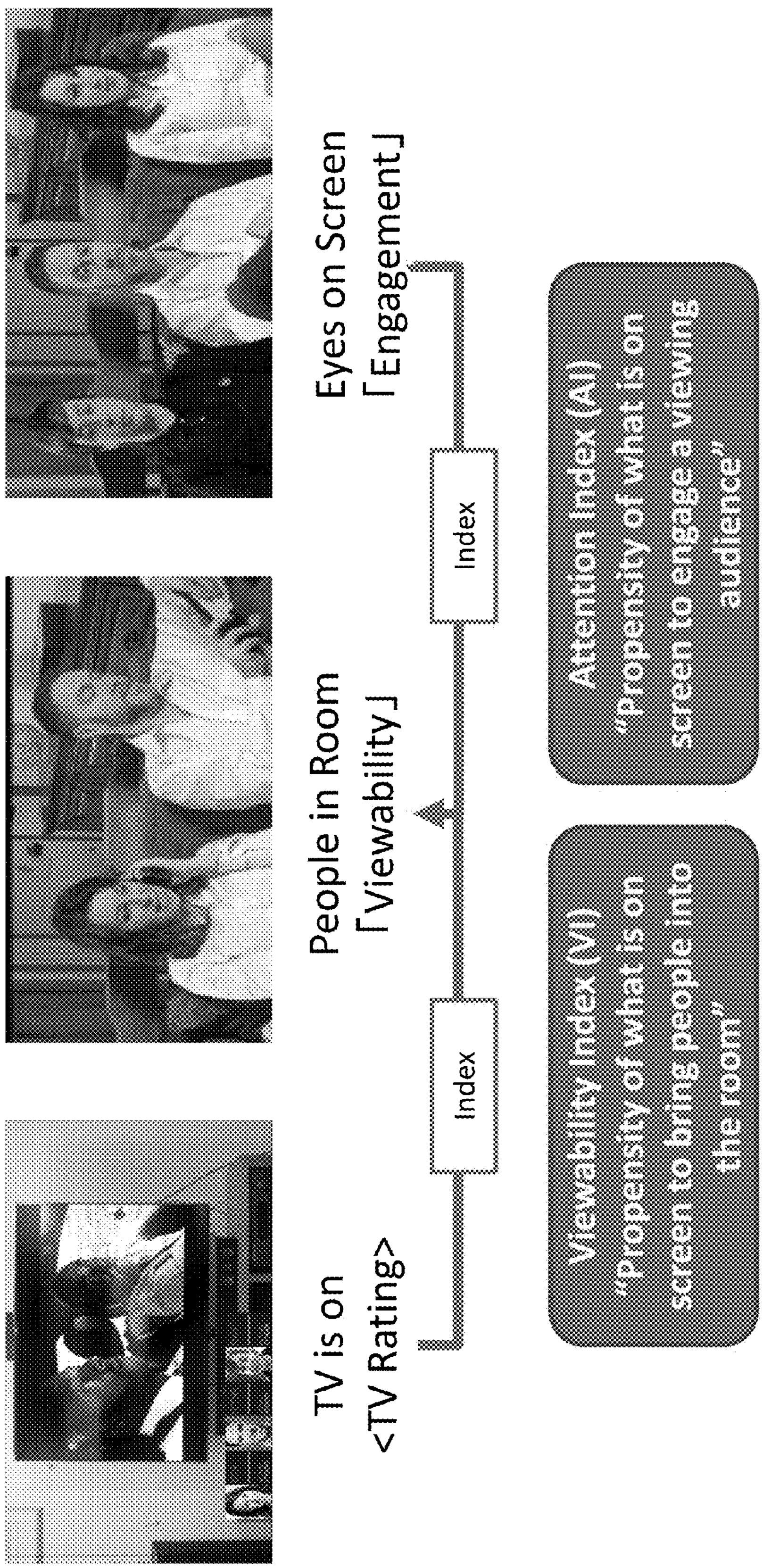

TV is on
<TV Rating>
People in Room
「Viewability」
Eyes on Screen
「Engagement」
Index
Index
Viewability Index (VI)
“Propensity of what is on screen to bring people into the room”
Attention Index (AI)
“Propensity of what is on screen to engage a viewing audience”

FIG. 3B

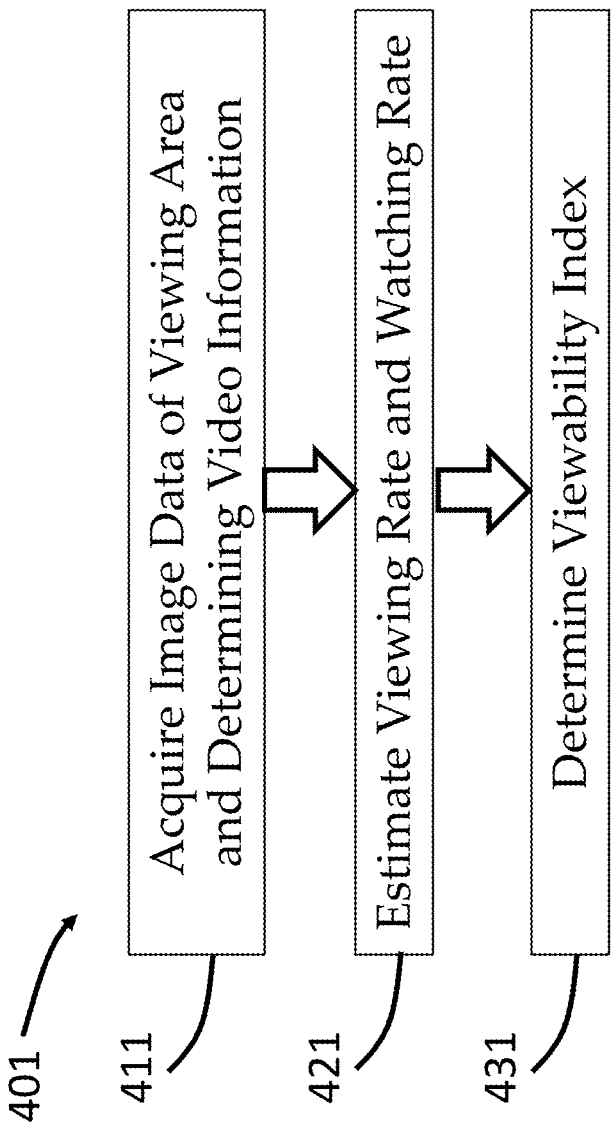

401
411
Acquire Image Data of Viewing Area and Determining Video Information
421
Estimate Viewing Rate and Watching Rate
431
Determine Viewability Index

FIG. 4A

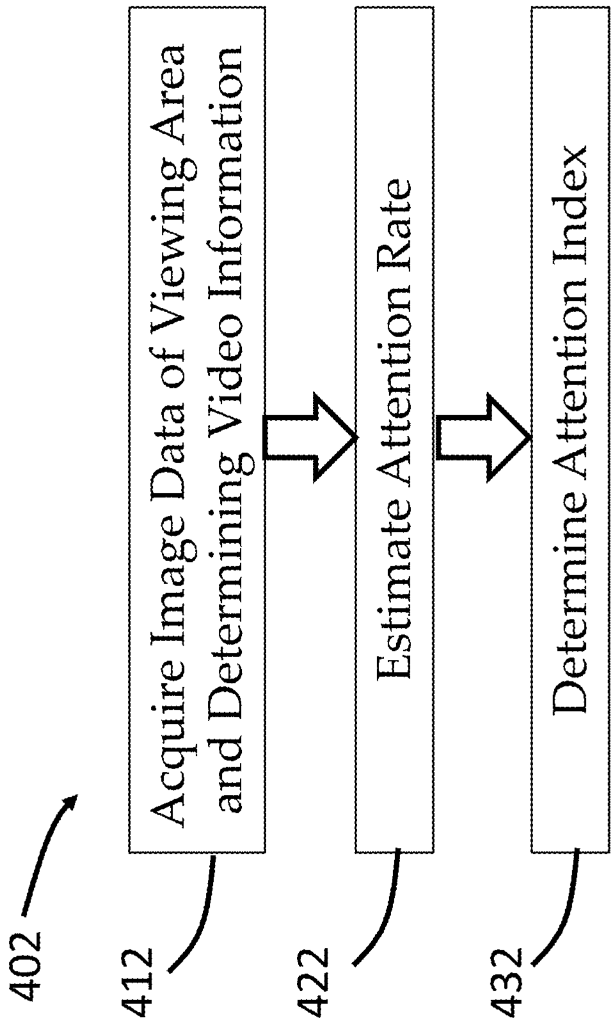

402
Acquire Image Data of Viewing Area and Determining Video Information
412
Estimate Attention Rate
422
Determine Attention Index
432

FIG. 4B

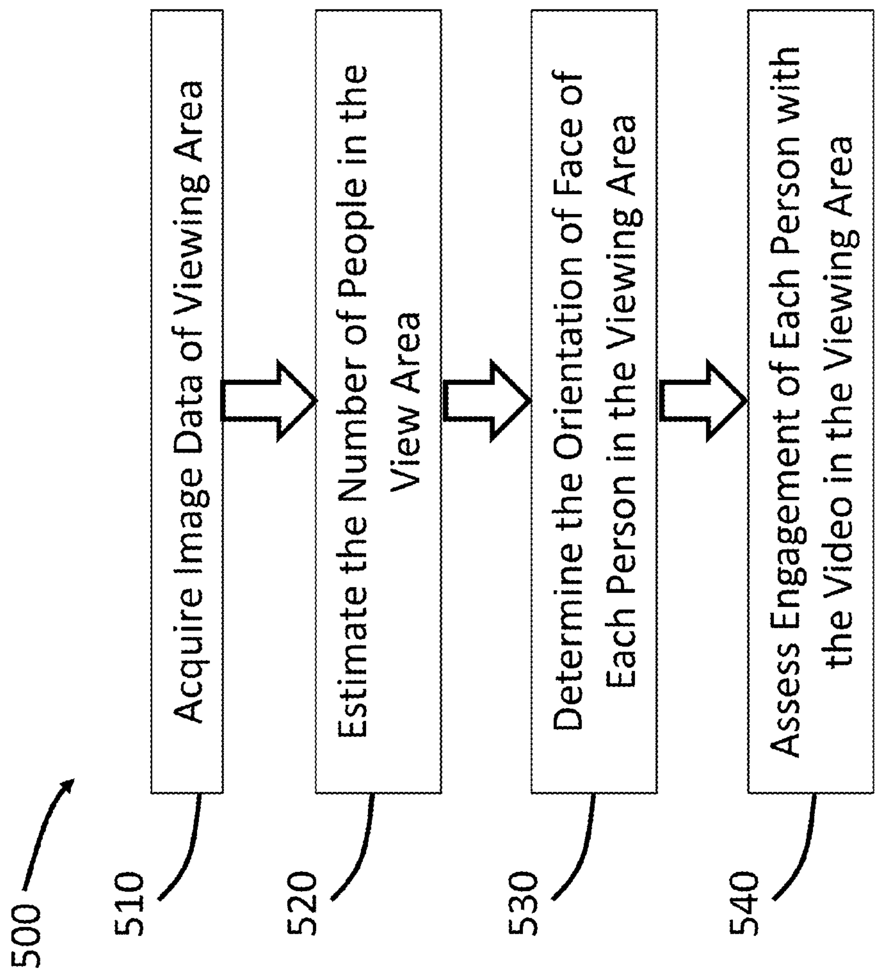

500
510
Acquire Image Data of Viewing Area
520
Estimate the Number of People in the View Area
530
Determine the Orientation of Face of Each Person in the Viewing Area
540
Assess Engagement of Each Person with the Video in the Viewing Area

FIG. 5

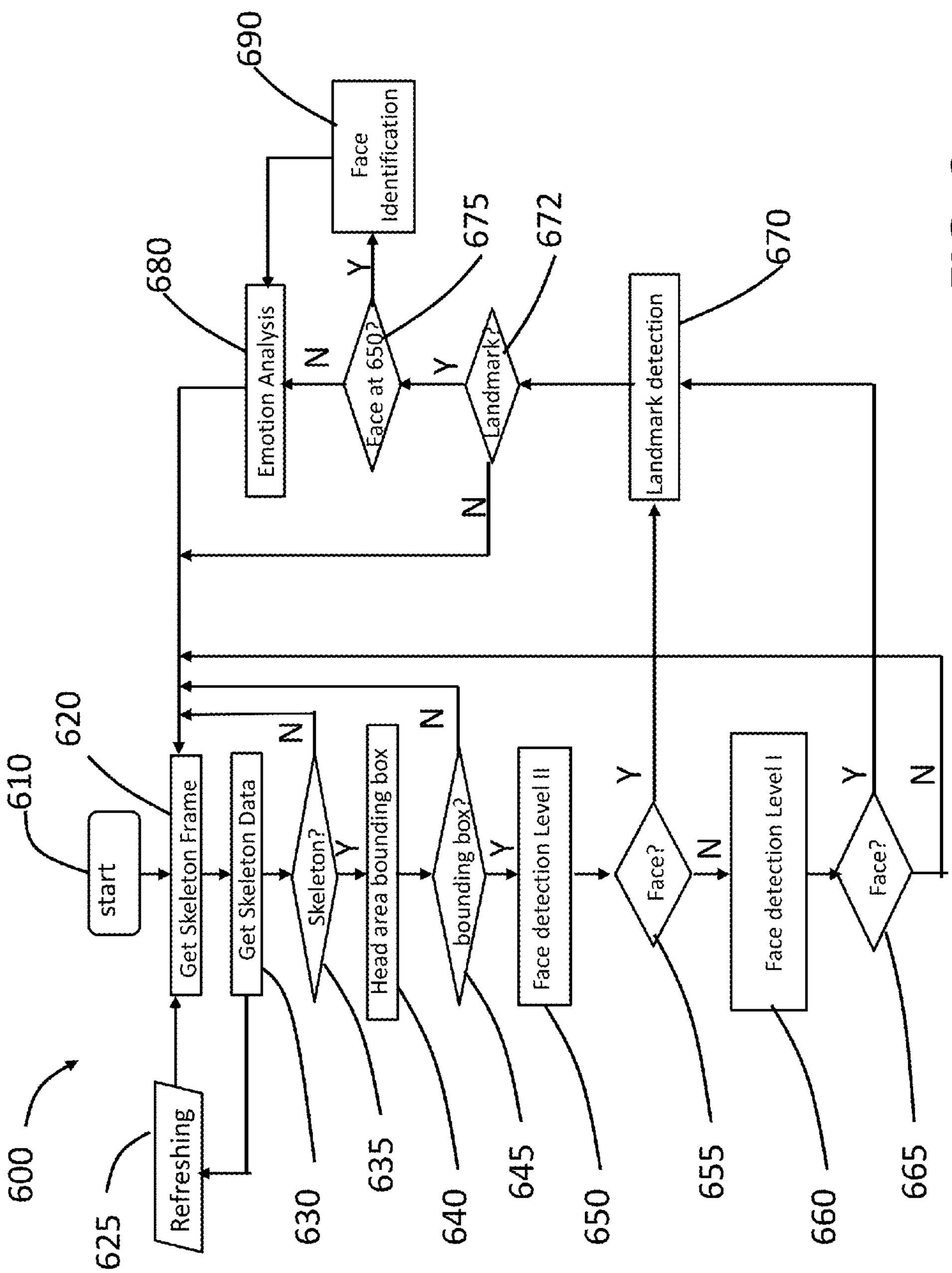

600
610
start
620
Get Skeleton Frame
625
Refreshing
Get Skeleton Data
630
Skeleton?
635
Y
N
Head area bounding box
640
bounding box?
645
Y
N
Face detection Level II
650
Face?
655
Y
N
Face detection Level I
660
Face?
665
Y
N
Landmark detection
670
Landmark?
672
Y
N
Face at 650?
675
Y
N
Emotion Analysis
680
Face Identification
690

FIG. 6

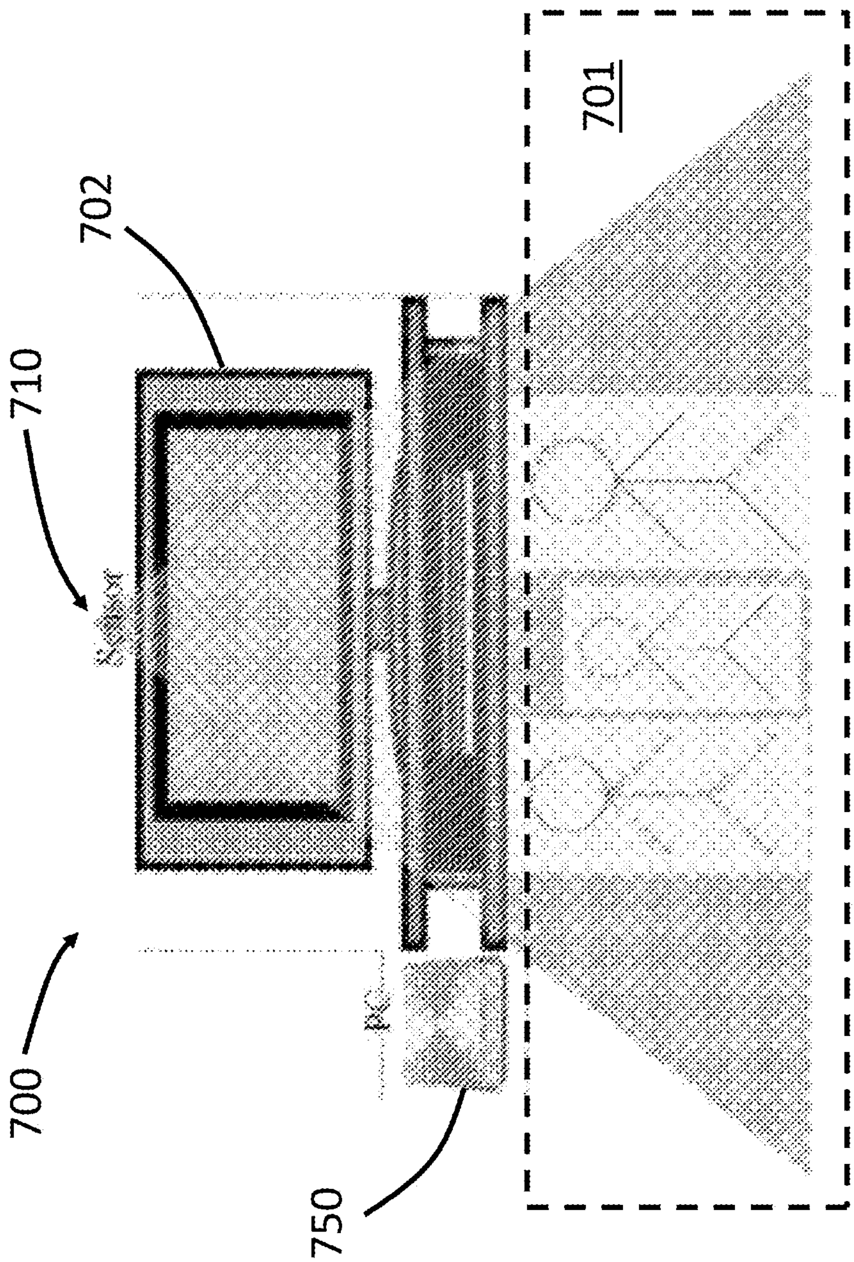


FIG. 7

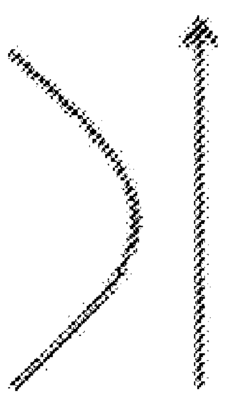

A. Downward convex

FIG. 8A

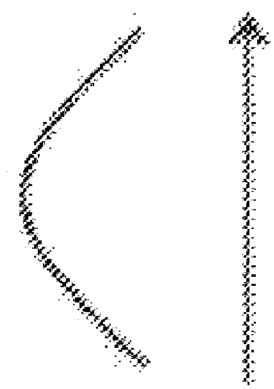

B. Upward convex

FIG. 8B

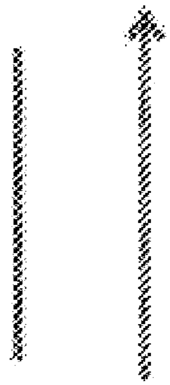

C. Parallel

FIG. 8C

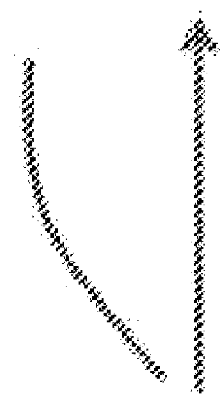

D. Increases Monotonically

FIG. 8D

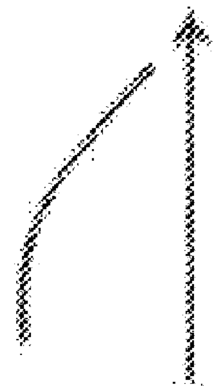
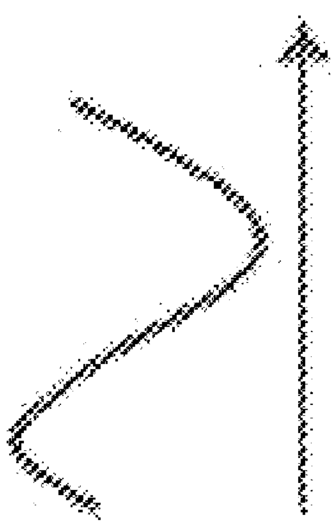
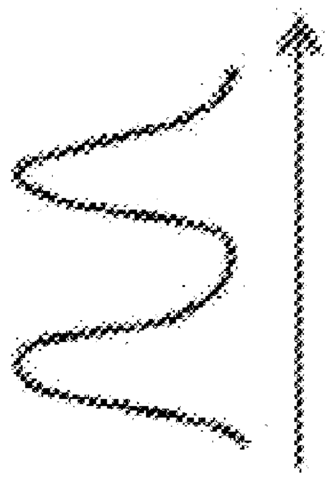

E. Decreases Monotonically

FIG. 8E



F. No regularity

FIG. 8F



G. Two mountains

FIG. 8G

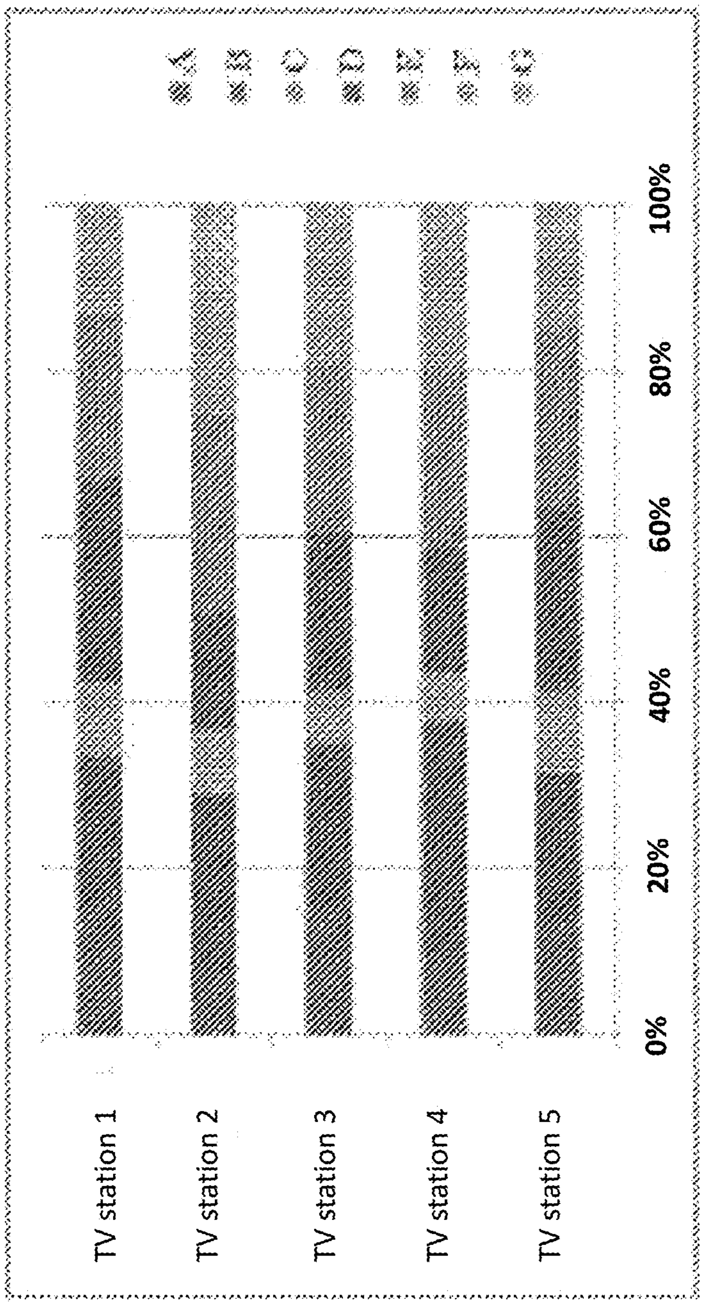

TV station 1
TV station 2
TV station 3
TV station 4
TV station 5
0%
20%
40%
60%
80%
100%

FIG. 9

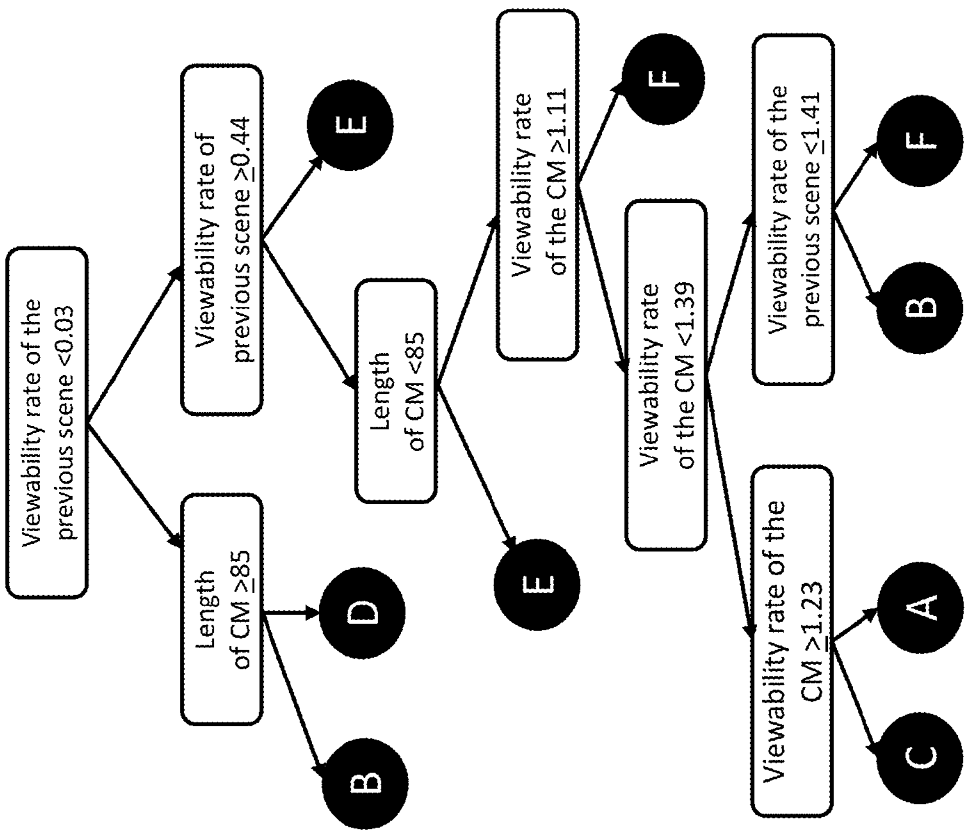

Viewability rate of the previous scene <0.03
Viewability rate of previous scene ≥0.44
Length of CM ≥85
E
Length of CM <85
D
B
Viewability rate of the CM ≥1.11
E
F
Viewability rate of the CM <1.39
Viewability rate of the previous scene ≤1.41
Viewability rate of the CM ≥1.23
F
B
A
C

FIG. 10

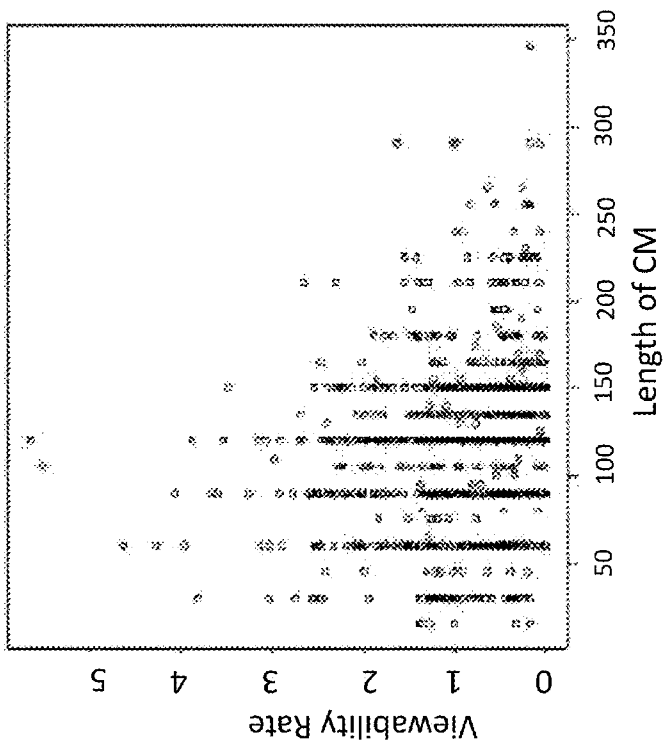

Viewability Rate
0
1
2
3
4
5
Length of CM
50
100
150
200
250
300
350

FIG. 11

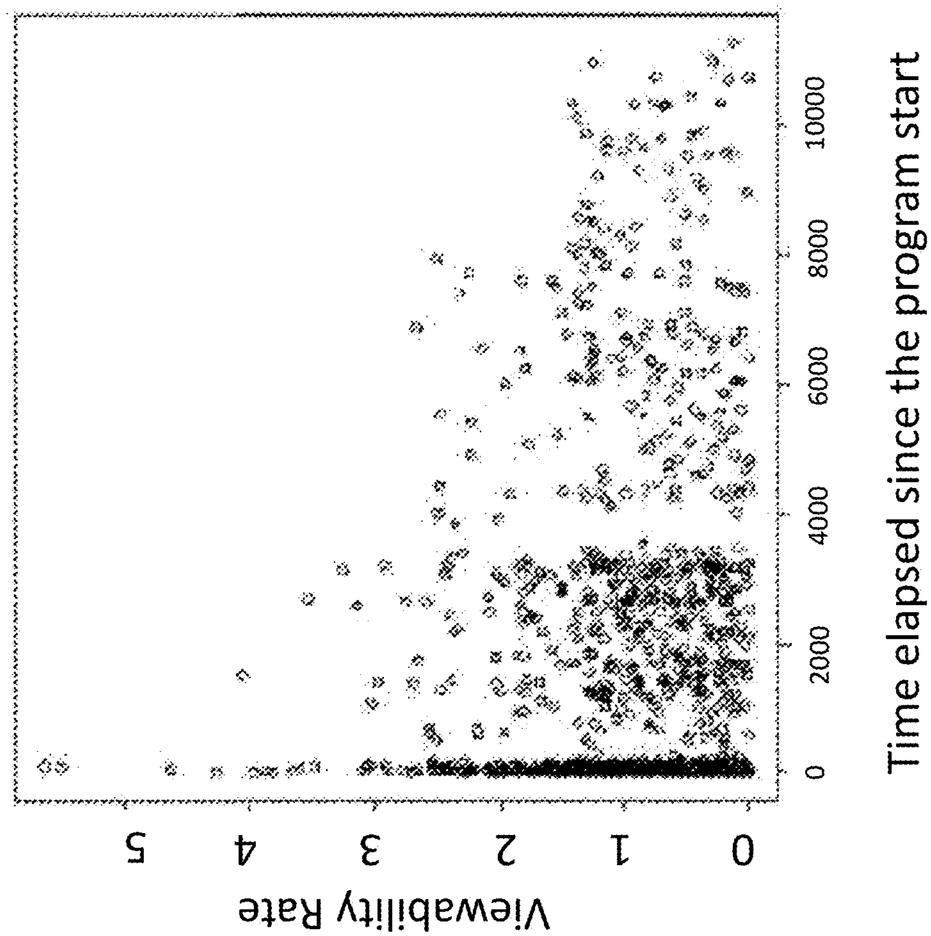

Viewability Rate
0
1
2
3
4
5
Time elapsed since the program start
0
2000
4000
6000
8000
10000

FIG. 12

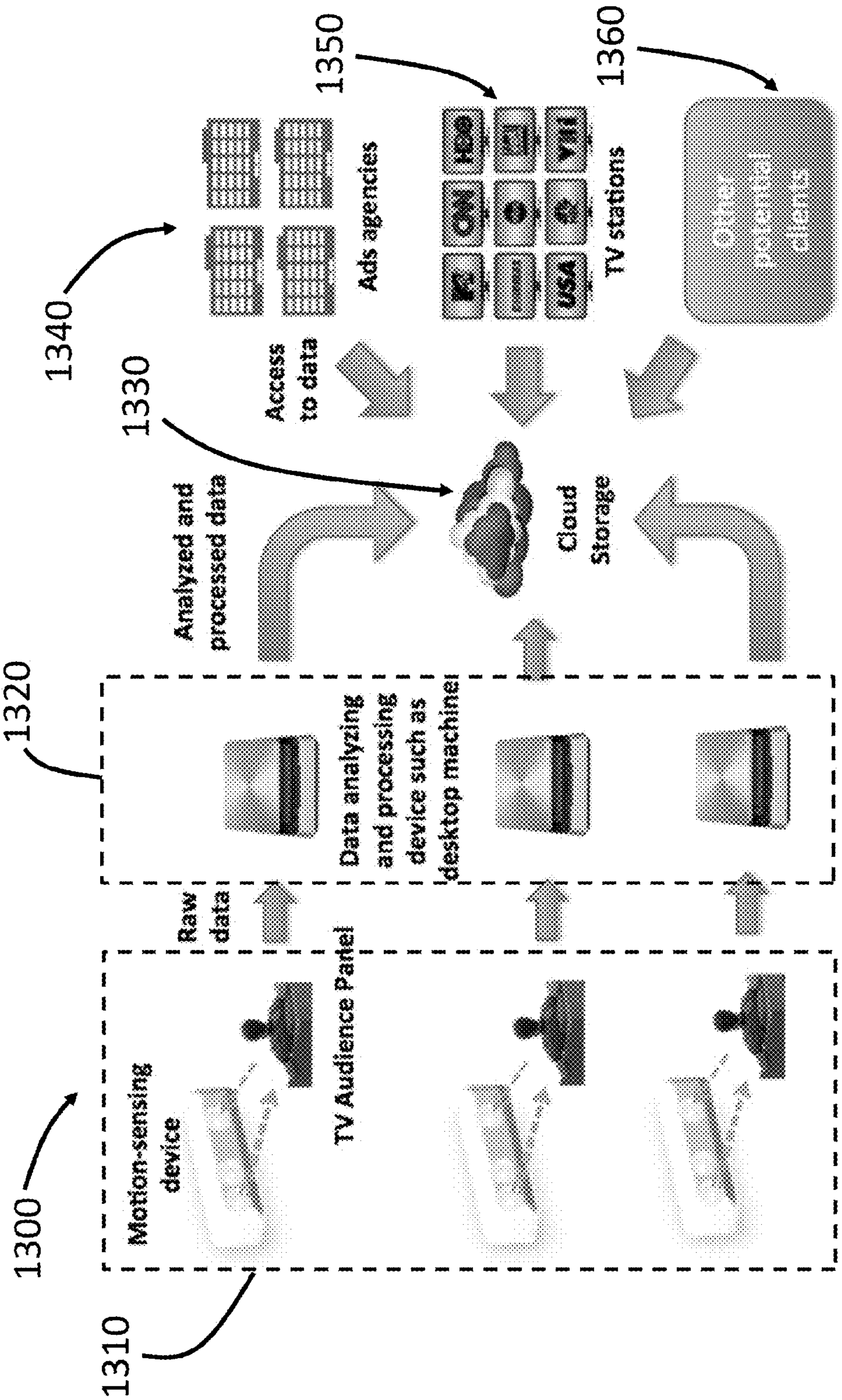

1300
1310
Motion-sensing device
TV Audience Panel
Raw data
1320
Data analyzing and processing device such as desktop machine
Analyzed and processed data
1330
Cloud Storage
Access to data
1340
Ads agencies
1350
TV stations
1360
Other potential clients

FIG. 13

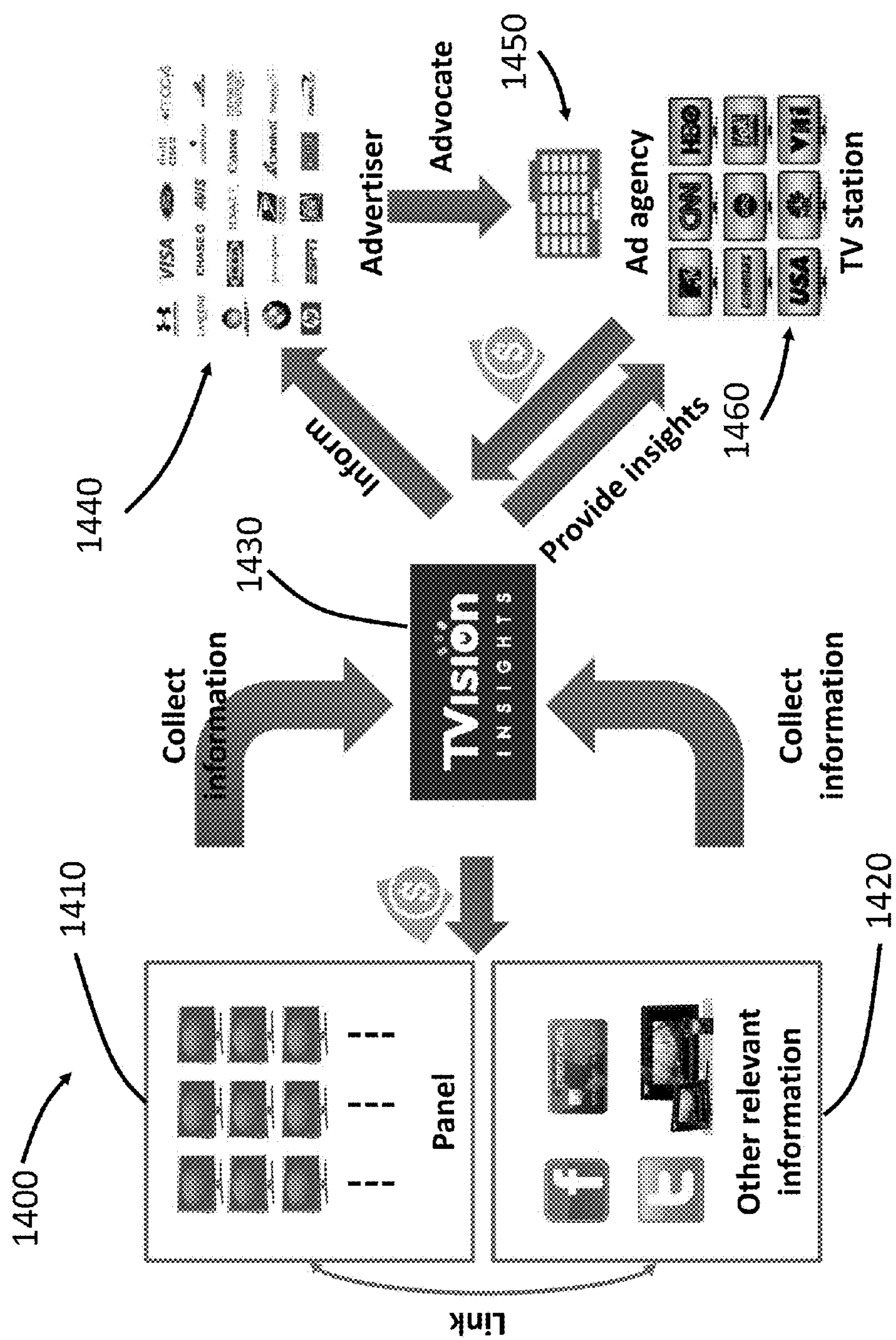

1400
1410
Panel
1420
Other relevant information
Link
Collect information
Collect information
TVision INSIGHTS
1430
1440
Inform
Advertiser
Advocate
1450
Ad agency
TV station
Provide insights
1460

FIG. 14

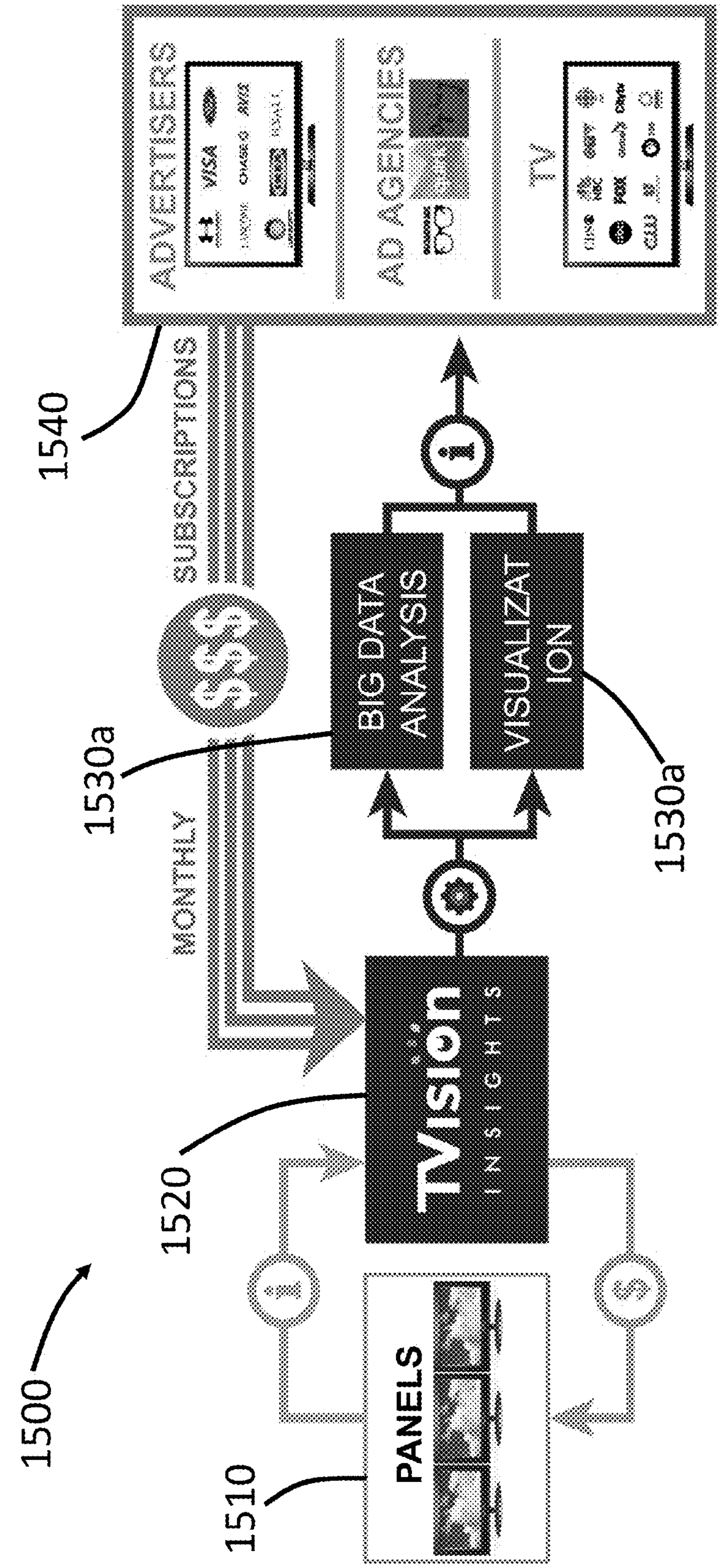

1500
1510
PANELS
1520
TVision
INSIGHTS
MONTHLY
SUBSCRIPTIONS
1530a
BIG DATA
ANALYSIS
VISUALIZAT
ION
1530a
1540
ADVERTISERS
AD AGENCIES
TV

FIG. 15

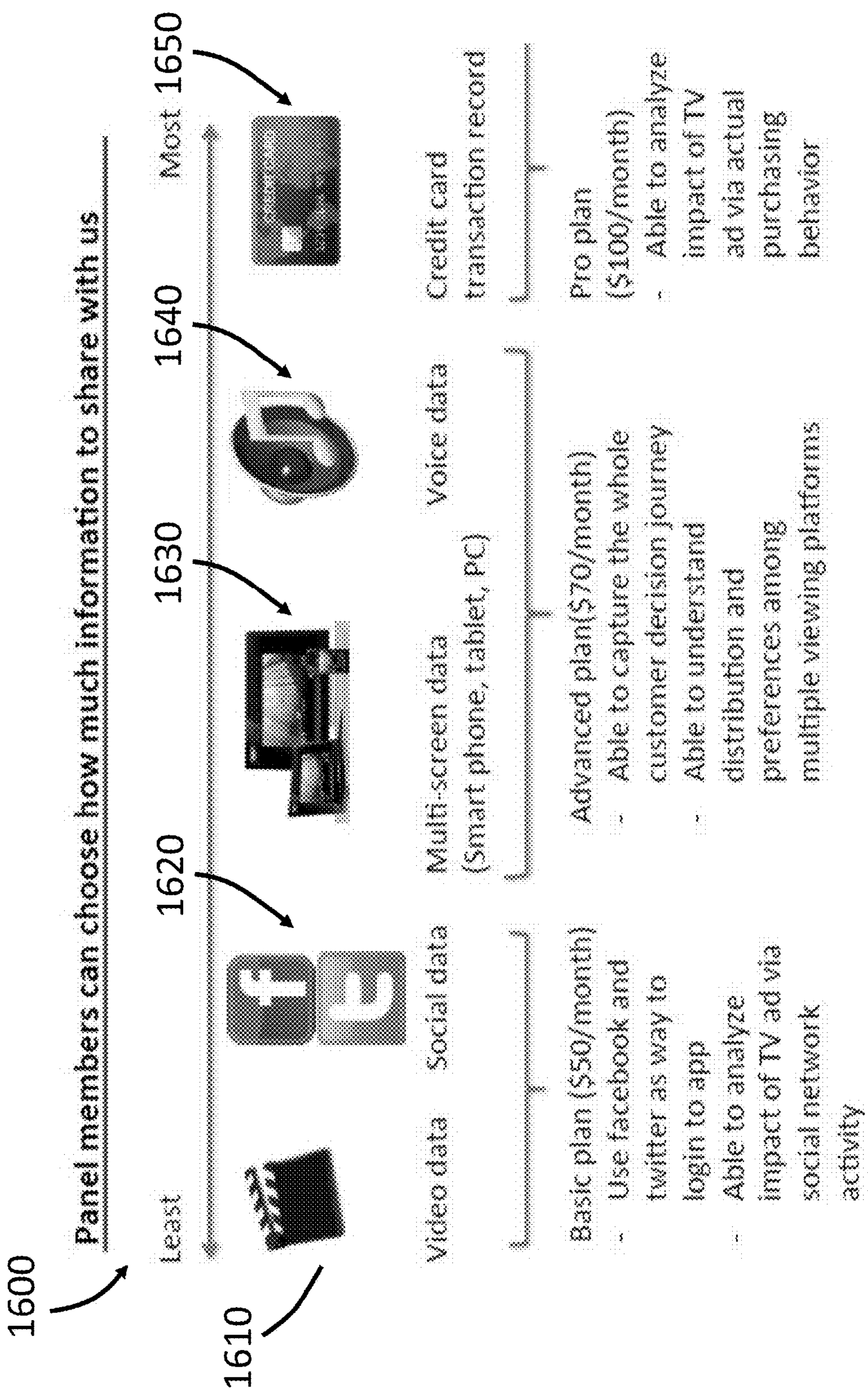

1600
Panel members can choose how much information to share with us
Least
Most
1610
1620
1630
1640
1650
Video data
Social data
Multi-screen data (Smart phone, tablet, PC)
Voice data
Credit card transaction record
Basic plan ($50/month)
- Use facebook and twitter as way to login to app
- Able to analyze impact of TV ad via social network activity
Advanced plan($70/month)
- Able to capture the whole customer decision journey
- Able to understand distribution and preferences among multiple viewing platforms
Pro plan ($100/month)
- Able to analyze impact of TV ad via actual purchasing behavior

FIG. 16

SYSTEMS AND METHODS FOR ASSESSING VIEWER ENGAGEMENT

CROSS-REFERENCE TO RELATED APPLICATION(S)

[0001] This application is a bypass continuation of PCT Application No. PCT/US2017/012531, filed Jan. 6, 2017, entitled "SYSTEMS AND METHODS FOR ASSESSING VIEWER ENGAGEMENT," which is hereby incorporated herein by reference in its entirety and which claims priority to U.S. Application No. 62/275,699, filed Jan. 6, 2016, entitled "SYSTEMS AND METHODS FOR ASSESSING VIEWER ENGAGEMENT," which is also hereby incorporated herein by reference in its entirety.

BACKGROUND

[0002] Conventional methods of TV audience measurements include using people meters and diaries to collect data from the audience. These methods typically try to recognize humans (potential audience members) in a room where a TV set is placed. The methods may also involve capturing a series of images (e.g., TV programs or commercial advertisements) playing on the TV. Then for each image, the number of people in the room at the time when a particular image is displayed can be estimated.

[0003] These methods have several flaws. First, the data collected by these methods normally only include the number of people in the room where the TV is placed. The data typically gives no indication of how often the viewer is actually watching the TV (the measurement takes place when the TV is on). Second, the collected data may indicate how often people are tuning to specific channels. However, it does not gauge their reaction to the programs or advertisements and therefore provides no indication of the effectiveness of the programs or advertisements. Third, TV ratings are not given for specific demographics in the household or in the community.

SUMMARY

[0004] Embodiments of the present invention include apparatus, systems, and methods of assessing viewer engagement of a TV audience. In one example, a system for quantifying viewer engagement with a video playing on a display includes at least one camera, disposed to image a viewing area in front of the display, to acquire image data of the viewing area. A microphone is disposed in proximity to the display to acquire audio data emitted by a speaker coupled to the display. The system also includes a memory, operably coupled to the camera and the microphone, to store processor-executable instructions and a processor, operably coupled to the camera, the microphone, and the memory. Upon execution of the processor-executable instructions, the processor receives the image data from the camera and the audio data from the microphone and determines an identity of the video displayed on the display based at least in part on the audio data. The processor also estimates, based at least in part on the image data, a first number of people present in the viewing area and a second number of people engaged with the video in the viewing area. The processor further quantifies the viewer engagement of the video based at least in part on the first number of people and the second number of people.

[0005] In another example, a method of quantifying viewer engagement with a video shown on a display includes acquiring, with at least one camera, images of a viewing area in front of the display while the video is being shown on the display. The method also includes acquiring, with a microphone, audio data representing a soundtrack of the video emitted by a speaker coupled to the display. The method further includes determining, with a processor operably coupled to the camera and the processor, an identity of the video based at least in part on the audio data and estimating, with the processor and based at least in part on the image data, a first number of people present in the viewing area while the video is being shown on the display and a second number of people engaged with the video in the viewing area. The method also includes transmitting, by the processor, the identity of the video, the first number of people, and the second number of people to a remote server.

[0006] In yet another example, a system for assessing viewer engagement with a video playing on a display is disclosed. The display is coupled to a speaker emitting a soundtrack of the video. The system includes a visible camera to acquire visible images of a viewing area in front of the display at a first sample rate while the video is playing on the display. An infrared camera is included in the system to acquire infrared images of the viewing area in front of the display while the video is playing on the display at the first sample rate. A microphone is disposed in proximity to the display to acquire samples of the soundtrack emitted by the speaker while the video is playing on the display at a second sample rate lower than the first sample rate. The system also includes a processor, operably coupled to the visible camera, the infrared camera, and the microphone, to: (i) identify the video based on the samples of the soundtrack, (ii) estimate, based on the visible images and the infrared images, a number of people in the viewing area while the video is playing on the display and a number of people engaged with the video, and (iii) overwrite, erase, and/or discard the samples of the soundtrack, the visible images, and the infrared images. The system also includes a memory, operably coupled to the processor, to store representations of an identity of the video, the number of people in the viewing area while the video is playing on the display, and the number of people engaged with the video. The system further includes a network interface, operably coupled to the processor, to transmit the representations to a server.

[0007] In yet another example, a method of quantifying viewer engagement for unique videos in a plurality of videos includes at each household in a plurality of households, acquiring image data of a viewing area in front of a display and determining if the display is showing a video in the plurality of videos. The method also includes, for each unique video in the plurality of videos, estimating (i) a viewing rate and (ii) a watching rate based on the image data and on demographic information about each household in the plurality of households. The viewing rate represents a ratio of a total number of people in the viewing areas to a total number of displays showing videos and the watching rate representing a ratio of a total number of people in households with display showing videos to a total number of people in the plurality of households. The method also includes, for each unique video in the plurality of videos, determining a viewability index based on the viewing rate and the watching rate.

[0008] It should be appreciated that all combinations of the foregoing concepts and additional concepts discussed in greater detail below (provided such concepts are not mutually inconsistent) are contemplated as being part of the inventive subject matter disclosed herein. In particular, all combinations of claimed subject matter appearing at the end of this disclosure are contemplated as being part of the inventive subject matter disclosed herein. It should also be appreciated that terminology explicitly employed herein that also may appear in any disclosure incorporated by reference should be accorded a meaning most consistent with the particular concepts disclosed herein.

BRIEF DESCRIPTION OF THE DRAWINGS

[0009] The skilled artisan will understand that the drawings primarily are for illustrative purposes and are not intended to limit the scope of the inventive subject matter described herein. The drawings are not necessarily to scale; in some instances, various aspects of the inventive subject matter disclosed herein may be shown exaggerated or enlarged in the drawings to facilitate an understanding of different features. In the drawings, like reference characters generally refer to like features (e.g., functionally similar and/or structurally similar elements).

[0010] FIG. **1** shows a schematic of a system for assessing viewer engagement of TV audiences.

[0011] FIG. **2**A illustrates a method of quantifying user engagement using the system shown in FIG. **1**.

[0012] FIG. **2**B illustrates a method of training a computer vision model for quantifying user engagement.

[0013] FIG. **3**A illustrates methods of viewer engagement including facial and eyeball tracking, facial recognition, and sentimental analysis.

[0014] FIG. **3**B illustrates the concepts of viewability index and attention index.

[0015] FIG. **4**A illustrates a process for assessing viewer engagement including estimating viewability index.

[0016] FIG. **4**B illustrates a process for assessing viewer engagement including estimating attention index.

[0017] FIG. **5** illustrates a process for assessing viewer engagement including determining the orientation of the face of each person in a viewing area.

[0018] FIG. **6** illustrates a process for detecting skeleton, face, identification, emotion, and engagement.

[0019] FIG. **7** shows a schematic view of a data acquisition architecture in exemplary methods of viewer engagement assessment.

[0020] FIGS. **8**A-**8**G show commercial message (CM) curves acquired using the architecture shown in FIG. **7**.

[0021] FIG. **9** shows the ratios of the CM curves for each of the sampled TV stations.

[0022] FIG. **10** shows a classification model through a decision tree with the determination results of the decision tree shown in TABLE 5.

[0023] FIG. **11** illustrates the viewability rate with respect to the length of the CM.

[0024] FIG. **12** shows the correlation between elapsed time since the start of the program and the viewability rate.

[0025] FIG. **13** illustrates communication of viewer engagement data acquired using the technology illustrated in FIGS. **1-12**.

[0026] FIG. **14** illustrates dissemination and use of viewer engagement data acquired using the technology illustrated in FIGS. **1-12**.

[0027] FIG. **15** illustrates big data analysis and visualization of viewer engagement data acquired using the technology illustrated in FIGS. **1-12**.

[0028] FIG. **16** shows a model for acquiring additional data to complement viewer engagement data acquired using the technology illustrated in FIGS. **1-12**.

DETAILED DESCRIPTION

[0029] To address shortcomings in conventional methods of TV audience measurements, systems and methods disclosed herein acquire image data of a viewing area in front of a display (e.g., a TV, computer, or tablet) that is playing a video (e.g., a TV show, movie, web show, advertisement, or other content). An example system determines how many people are in the viewing area and which of those people are actually watching the video from the image data. The system also samples the soundtrack of the video with a microphone and identifies the videos using the samples of the soundtrack. The system stores (and/or persists) information about the video, the number of people in the viewing area, and the number of people watching the video in a local memory and transmits the information to a remote server via an internet or other network connection.

[0030] Unlike previous systems for measuring viewer engagement with videos, which identify videos based on digital watermarks embedded in the videos themselves, examples of the inventive system identify videos based on the videos' soundtracks. As a result, the inventive systems do not have to be connected to the display, the set-top box, or the cable connection at the viewer's premises. This makes them easier to install and remove (and thus more likely to be adopted). It also makes them less likely to malfunction or to record "false positive" impressions caused by leaving the set-top box on while the display is off.

[0031] An inventive system also processes image data locally, i.e., on the viewer's premises, to determine the numbers of people in the viewing area and engaged with the video. It can also process audio data locally to identify the video being displayed while someone is in the viewing area. It stores this data locally, i.e., in a memory in or coupled to a local device on the viewer's premises. The processed image and audio data consumes far less memory than the raw image and audio data, so this local memory can store information covering longer time periods. In other words, an inventive device uses memory more efficiently because it stores processed data instead of raw data.

[0032] The local device processes the raw image data, which may include both visual and depth information, acquired from the viewing area to assess viewer engagement. The local device can use artificial intelligence (AI) technology and machine learning techniques to analyze a viewer's body gestures, movements, and facial orientation. The local device can also recognize individual faces in the video audience and determine each viewer's emotions from the image data. In this processing, the individual's images are not transmitted outside of the individual's premises. The recognition can be performed on the local device on premises. Each individual in the household can receive a unique identifier during the on-boarding process for that household. When a match is made during the recognition process, this identifier is assigned to the match, and this identifier can then be transmitted to remote servers. In addition, the processing is carried out over the streaming video or audio

data (including images). In other words, the video or audio data is not persisted in local memory.

[0033] The local device processes the raw audio data by matching or comparing the raw audio data with samples in an audio database to identify the specific video (e.g., TV channel, program, or advertisement) that is being viewed. Alternatively, or additionally, the local device can submit a query based on the audio data to a third-party application programming interface (API), which identifies and returns an identification of the content to which the audio belongs. In some cases, the database or API may return multiple possible matches, and the remote server can select the best match using information about the TV schedule, subsequent audio samples, or data collected from other sources, including, but not limited to, the set-top box, cable/internet connection, or the content provider itself.

[0034] In some implementations, the local device does not store the raw image or audio data for later retrieval. Instead, the local device writes the raw image and audio data to one or more buffers that store the raw image and audio data for processing, then overwrites or erases the buffers after the raw image and audio data has been processed. Put differently, the local device holds the raw image and audio data merely transiently during processing. As used herein, "holding" of raw images and audio data in local devices refers to temporary storing of these data for a short time duration (e.g., less than 100 milliseconds, less than 80 milliseconds, less than 60 milliseconds, less than 50 milliseconds, or less than 40 milliseconds, including any values and sub ranges in between). Overwriting or erasing the raw image and audio data offers a number of advantages, including reducing the amount of memory required by the local device. It also enables easier compliance with data privacy laws by eliminating image or audio data that could be used to identify people, including children, in the viewing area or in range of the microphone.

[0035] Processing and storing image and audio data locally offers another technical advantage—it reduces the bandwidth required to convey information about viewing habits from the local device to the remote server. Compared to raw image and audio data, processed image and audio data consumes less memory and therefore requires less bandwidth for transmission. The processed image and audio data also fills a given memory more slowly than raw image and audio data and therefore can be transmitted to the remote server less frequently. A local device may take advantage of this flexibility by scheduling burst transmissions during times when network bandwidth usage is relatively low, e.g., late night or early morning. Transmitting processed image and audio data, which doesn't necessarily include information identifying people, including children, in the viewing area or in range of the microphone, also ensures or increases the ease of compliance with data privacy laws.

[0036] The remote server collects processed image and audio data from local devices in different households. It processes this data to assess viewer engagement across an entire community by statistically analyzing the viewer engagement information collected from the different households in the community. For example, the server can quantify the ratio of the viewer engagement from the highly granular data collected from each household to the total length of the programming that was detected.

[0037] The statistical analysis can further take into account demographic information (e.g., age, gender, household income, ethnicity, etc.) of the people watching the videos and/or the people in the household. Based on all this information, the server may calculate various indices, such as a viewability index and an attention index (both defined below), to quantify viewer engagement. These viewer engagement indices may be based on any and all information provided by the local devices, including information about the viewers' body gesture(s), movement(s), and facial orientation(s) of viewers, as well as the video information. These quantitative indices can indicate, among other things, (i) who is/are really watching display, (ii) how often an audience member looks at the display, and ii) the audience's reaction towards the programs and advertisements on the display.

[0038] The quantitative indices can then be transferred by the remote server to a central storage (e.g., a cloud-based database) where third parties, including but not limited to TV advertising agencies and TV networks, can access the indices and possibly other data as well. Alternatively, the raw data collected by the sensors can be transferred to a central storage on the cloud where it is analyzed by methods described herein and made available to interested third parties. A third party may optionally access the raw data through the system. The raw data in this example includes data collected after processing of the video and audio streams (instead of the video and audio streams themselves). Generally, speaking, the raw data can include unique identifiers of the viewers, the attentiveness of the viewer(s), and the programming being viewed by the viewer(s), on a sub second basis (e.g., every half second or less). More quantitative indices (see more details below) can be computed on the remote server using this raw data.

[0039] This acquired and analyzed data can allow a collection entity, such as a content provider or advertising agency, to accurately evaluate the impact of videos, including unprecedented measurements of individual demographics, which can be valuable to the advertisers. For example, advertising agencies can use the data to determine which commercial slots would be a best fit for their targeted audience. With demographic information, the data can be matched to the type of audience and can effectively lead to purchasing behavior, thereby increasing return on investment (ROI) in programming. TV networks can also benefit from the data as they can glean more accurate ratings of their TV programs, audience type, reactions, and predictive ad slot value. This further allows them to improve their programs to better fit the type of audience and eliminate less popular shows, in addition to determining which ad slots may have the highest value for a particular target demographic.

[0040] The acquired and analyzed data also allows various business models. For example, a collection entity can provide performance-based TV ratings data and raw data for analysis, which is collected from a motion-sensing device put into selected-user households that represent a national and/or local demographic, to TV networks, advertising agencies, and other interested third parties and indirectly to advertisers who obtain the data from advertising agencies.

[0041] Systems of Assessing Viewer Engagement

[0042] FIG. **1** illustrates a schematic view of a system **100** for assessing viewer engagement in a household, a sports bar, or other space with a display. The system **100** includes a local device **105** disposed in each household to collect viewer engagement data and a remote server **170**, such as a

cloud storage and computing device that includes a memory to store data and a processor (also called a remote processor) to analyze data. The local device **105** is communicatively coupled to the remote server **170** via a network connection **172**, such as an internet connection. For instance, the local device **105** may include a network interface **165**, such as a WiFi antenna or Ethernet port, for connecting to a household local area network (LAN). This LAN is in turn connected to a wide area network (WAN), e.g., via a cable or fiber optic connection provided by an Internet Service Provider (ISP).

[0043] The local device **105** in FIG. **1** includes an infrared (IR) emitter **110** to illuminate a viewing area **101** in front of a display **11**, such as a television (TV), computer screen, tablet, or other device, with IR light. This IR light may be structured or modulated to produce an illumination pattern that scatters or reflects off objects (including the human audience) in the viewing area **101**. The local device **105** also includes an IR sensor **120** that detects the IR light reflected or scattered by these objects. A processor **150** (also called a local processor **150**) coupled to the IR emitter **110** and IR sensor **120** uses information about the illumination pattern and the detected IR light to produce one or more IR depth images or IR depth maps of the viewing area **101**. More specifically, the processor **150** converts information derived from the reflected beams into depth information measuring the distance between a viewer and the sensor **120**. The processor **150** uses these IR depth images to determine how many people are in the viewing area and which of those people are watching the display. The processor **150** may also derive information from the IR depth images about the identities of the people watching the display, possibly by recognizing their faces or gestures or determining their demographics (e.g., age, gender, etc.).

[0044] The local device **105** further includes an RGB sensor **130** (also referred to as a visible camera) that captures color images of the viewing area **101**. The processor **150** is also coupled to the RGB sensor and may use the color images, alone or in combination with the IR depth images, to estimate the number of people are in the viewing area, the number of people engaged with the display, and information about the people in the viewing area. The color images can also be used for facial recognition. In some cases, the processor **150** uses both the color images and the IR depth images to improve the fidelity of the estimates of the numbers of people in the viewing area and engaged with the video.

[0045] The local device **105** also includes one or more microphones **140** positioned to detect sound emitted by a speaker **13** coupled to the display **11**. In operation, the speaker **13** plays the soundtrack of the video shown on the display **11**. And the microphone **140** captures audio samples of the soundtrack played by the speaker **13**. The processor **150**, which is coupled to the microphone **140**, uses these audio samples to create an audio fingerprint of the video (soundtrack), which it compares with other audio fingerprints in a proprietary or third-party database to identify the video being shown on the display **11**.

[0046] The system **100** can further include a Bluetooth receiver **180** matched with a Bluetooth transmitter **185**. In some cases, the Bluetooth transmitter **185** can be included in a wristband or a wristwatch worn by the viewer. In operation, the Bluetooth transmitter **185** transmits a low power Bluetooth beacon, which is received by the Bluetooth receiver **180**. The processor **150** can then gauge the viewer's distance from the display **11** based on the received Bluetooth beacon. In addition, each Bluetooth transmitter **185** can have a unique ID that can be recognized by the processor **150**. The transmitter ID can be further associated with a unique viewer (e.g., each viewer in the household has his or her own transmitter). In this manner, the identity of the viewer can also be determined.

[0047] In some cases, the system **100** can include more than one Bluetooth receiver. These receivers can be disposed at different locations such that each receiver can receive different Bluetooth signal strength from the transmitter **185**. This configuration can allow the processor **150** to estimate not only the distance of the viewer from the display **11** but also the relative location of the viewer (e.g., to the left or right of the display **11**).

[0048] The system **100** may include other motion-sensing devices, such as a 3-axis accelerometer to detect position and motion. The motion-sensing device can be connected, for example, via a USB cable with a data-analyzing and processing device such as a desktop machine.

[0049] FIG. **1** shows the data collection components—here, the IR emitter **110**, IR sensor **120**, RGB sensor **130**, and microphone **140**—as part of the local device **105** (e.g., within the same housing). In other embodiments, one or more of these components may be implemented as separate devices that are coupled to the processor **150** by one or more wired connections, such as USB connections, RS 232 connections, Ethernet connections, fiber connections, or one or more wireless connections, such as WiFi connections, Bluetooth connections, other RF connections, or infrared connections. For instance, the IR emitter **110** and IR sensor **120** may be (in) a commercially available device, such as a Microsoft Kinect, that is connected to the processor **150**. Likewise, the microphone **140** may be implemented as an array of microphones that are placed around the viewing area or close to the speaker **13**. A microphone array may be better able to extract voice input from ambient noises. The local device **105** may include or be coupled to other sensors as well.

[0050] The processer **150** in the system **100** is employed to process the raw data acquired by the sensors, including the IR emitter **110**, the IR sensor **120**, the RGB sensor **130**, and the microphone **140**. The processing can be carried out upon execution of processor-executable instructions that are stored in a memory **160** coupled to the processor **150**. In one example, a user can manually store the instructions in the memory **160** by downloading the instructions from the remote server **170**. In another example, the local device **105** can be configured to (routinely) check whether there are updated instructions available for downloading from the remote server **170**. If so, the local device **105** can automatically download the update via the network connection **172** and the network interface **165**. In yet another example, the remote server **170** can be configured to send a notification to the local device **105** when an update or a set of new instructions is ready for downloading. Upon receiving the notification, a user can decide whether to download and/or install the update. In yet another example, the remote server **170** can be configured to send update notification to another user device, such as a smartphone. Upon receiving the notification, the user can decide whether the download and/or install the update.

[0051] The memory **160** in the local device **105** also stores the processed data (e.g., the estimate of the number of

people in the viewing area, the estimate of the number of people engaged with the display, and the identification of the video, as well as any demographic information or indices derived from the raw image and audio data). Once the memory **160** has accumulated enough processed data, the processor **150** transmits the processed data to the remote server **170** via the network interface **165** and the network connection **172** for aggregation, further processing, and reporting. The local memory **160** also temporarily holds the image and audio data during the local processing. In some cases, this processing is completed in less than a quarter of a second.

[0052] Collecting and Processing Image and Audio Data with a Local Device

[0053] FIG. **2**A illustrates a process **200** for collecting and processing image and audio data acquired with a system like the system **100** shown in FIG. **1**. As described above, the system can include a visible sensor, an IR sensor, or both to images of the viewing area in front of the display (**202**). In one example, the RGB sensor **130** and the IR sensor **120** operate independently from each other; the sensors acquire images in an unsynchronized fashion. In another example, the image acquisition by the RGB sensor **130** and the IR sensor **120** is substantially synchronized. Each time the RGB sensor **130** acquires a visible image, the IR sensor **120** acquires an IR image, e.g., at the same time or in an interleaved fashion.

[0054] A local processor (e.g., processor **150**) detects the number of people in the images of the viewing area (**204**) and also determines which of those people are engaged with the display (**206**). For instance, the local processor may use the techniques described below, including skeleton detection techniques, facial recognition techniques, and eye tracking techniques as known in the art of computer vision/image processing. In some cases, the local processor **150** can determine additional indices related to the duration of each viewer's presence in the viewing area, the duration of each viewer's engagement with the display, and the identity of the video being displayed (**208**), which can be derived from audio data as described below (**222**).

[0055] The local processor can further identify each person detected in the viewing area **101**, on a demographic level (e.g., man aged 25-30, girl aged 12-15) (**210**). If the local processor **150** has access to information about the household where the local device **105** is placed, e.g., via the local memory **160** or the remote server **170**, it may use this demographic information to provide more confident demographic information estimates of each person detected in the viewing area **101**. The local processor may even identify the particular people in the household who are in the viewing area.

[0056] The local processor **150** can also estimate the mood or emotion of each person detected in the viewing area **101** (**212**). The emotions that can be determined by the processor **150** can include, for example, happy, sad, or neutral. The classification of a viewer's emotion, when watching a video on the display **11**, can be used to gauge the viewer's reaction to the video, thereby facilitating targeted delivery of advertisement.

[0057] To estimate the mood or emotion of each person, the local processor **150** can capture the visual information (e.g., from the images of the viewing area **101**) in real-time from both RGB and IR channels. The visual information can be further processed to extract patterns and features that can be signatures of different mood or emotion states. The features extracted from both channels can be fused as a unified feature. A classifier can be trained to take such feature as input. Estimation of emotion/mood can be then made based on the classifier/s response to certain patterns in each time.

[0058] In some cases, the estimation of mood or emotion can be achieved by the following method. The method includes collecting training images with people displaying various emotions, such as, smiling and frowning, among others. Features representative of each emotion are extracted (e.g., by a processor) from these training images. The features and the images are then used to train a classifier to correlate each feature to a corresponding emotion. In this manner, the classifier can assign these features to the various emotions. The method also includes deploying the classifier on the local device so as to recognize the viewers emotions in real time.

[0059] In cases where the system collects visible and IR images in a synchronized fashion, the visible and IR cameras can collect images for training a computer vision model used by the processor to detect people (**204**), count engaged viewers (**206**), identify viewers demographically (**210**), and estimate mood (**212**). The training can be employed to establish a "ground truth." Having collected image data from both IR and RGB sensors almost in parallel, a human can annotate the people detected in each image. This manual data can be fed to a training algorithm, giving rise to two separate models, one trained on visible RGB spectrum, and the other on the IR spectrum. The detection rate of each model against the "ground truth" is then compared to select the model that performs better. More details of this training are described below with reference to FIG. **2**B.

[0060] Synchronization of the two cameras (e.g., sensors **120** and **130** in FIG. **1**) can also allow the local processor to double-check the image processing. For example, the processor **150** can compare the number of people identified in each image or remove errors visible in one image and less visible or invisible in the other image. If the results are in agreement with each other, the processor **150** can record the results. If not, the processor **150** can then detect possible errors in at least one of the images. Alternatively, the processor **150** can generate an alert for a human to intervene. The processor **150** can also generate a flag associated with the data estimated from these two images, indicating that there this data might be less reliable. In subsequent analysis, this data may not be used at all, if images take shortly before or after this pair of images at issue can provide reliable people recognition.

[0061] In one example, the local device **105** uses the visible and IR sensors **120** and **130** all the time to take image data. In another example, the local device **105** can use only one of the sensors **120** or **130** to take image data. In yet another example, the local device **105** can use one sensor as a default sensor and use the other sensor as a backup sensor. For example, the local device **105** can use the RGB sensor **130** most of the time for image taking. However, if the processor **150** has trouble satisfactorily analyzing the visible images (e.g., the analysis is not as reliable as desired), the processor **150** can turn on the IR sensor **120** as backup (or vice versa). This may occur, for example, when the ambient light level in the viewing area is low.

[0062] The local processor may also adjust the image acquisition rate for the visible sensor, the IR sensor, or both

based on the number of people in viewing area, their positions in the viewing area, and the identity of the video on the display (**214**). Generally, the image acquisition for either or both sensors can be substantially equal to or greater than about 15 frames per second (fps) (e.g., about 15 fps, about 20 fps, about 30 fps, about 50 fps or even greater, including any values and sub ranges in between). At this image acquisition rate, the sensor can detect eye movements well enough for the local processor to assess viewer engagement (**206**).

[0063] The local processor may increase or decrease the image acquisition rate based on the number of people in the viewing area **101**. For example, if the processor determines that nobody is in the viewing area **101**, it may reduce the image acquisition rate to reduce power and memory consumption. Likewise, if the processor determines that the viewer(s) are not engaged with the video (e.g., because they appear to be sleeping), it may reduce the image acquisition rate to conserve power, memory, or both. Conversely, the processor may increase the image acquisition rate (e.g., to greater than 15 fps) if the viewers appear to be shifting their attention rapidly, if they are watching a fast-paced video (e.g., a football game or action movie), if they are changing channels rapidly (e.g., channel surfing), or if the content is changing relatively rapidly (e.g., during a series of advertisements).

[0064] If the system includes both IR and visible image sensors, the local processor may also vary the image acquisition based on the lighting conditions or relative image quality. For instance, in low light conditions, the local processor may acquire IR images at a higher rate than visible images. Similarly, if the local processor gets better results processing visible images than IR images, it may acquire visible images at a higher rate than IR images (or vice versa if the opposite is true).

[0065] The system also records samples of the video's soundtrack with the microphone **140** (**220**). Generally, the audio data acquisition rate or audio sampling rate is lower the image acquisition rate. For instance, the microphone acquires audio samples at a rate of once every 30 seconds. In each acquisition, the microphone **140** records an audio sample having a finite duration so as to allow identification of the video associated with the audio sample. The duration of the audio sample can be substantially equal to or greater than 5 seconds (e.g., about 5 seconds, about 6 seconds, about 8 seconds, about 10 seconds, about 20 seconds, or about 30 seconds, including any values and sub ranges in between).

[0066] The local processor uses the audio samples recorded by the microphone **140** to identify the video being played on the display (**222**). For example, the processor **150** can create a fingerprint of the audio data and use the fingerprint to query a third-party application programming interface (API), which responds to the query with an identification of the video associated with the audio data. In another example, the processor **150** can compare the fingerprint against a local table or memory to determine the identity of the video.

[0067] As mentioned above, using samples of the video soundtrack to identify the video offers several advantages over the digital watermarks used by conventional TV survey devices to identify videos. It does not require inserting digital watermarks into the video, which eliminates the need to coordinate with content producers and providers. This simplifies content production and distribution and makes it possible to identify and assess a wider range of video content, including producers and distributors who cannot or will not provide digital watermarks. And it eliminates the need to connect the local device to the cable or set-top box.

[0068] In addition, using audio data instead of digital watermarks reduces the risk of "false positives," or instances where the system detects people in the viewing area and identifies a video that is not actually being watched even when the TV is off. This can happen with a conventional system hooked to set-top box if the household members leave their set-top box on even when their TV is off.

[0069] In some examples, the local processor adjusts the audio sampling rate (**224**), e.g., based on the identity of the video, the number of people in the viewing area, the number of people engaged with the video, etc. For instance, if the local processor cannot identify the video from a single fingerprint (e.g., because the video soundtrack includes a popular song that appears in many different video soundtracks), the local processor and microphone may acquire samples at a higher rate or of longer duration to improve video resolve any ambiguity. The processor may also decrease the audio sampling rate if nobody is in the viewing area **101** or the viewer(s) are not engaged with the video (e.g., because they appear to be sleeping) to conserve power, memory, or both. Conversely, the processor may increase the audio sampling rate if the viewers are changing channels rapidly (e.g., channel surfing) or if the content is changing relatively rapidly (e.g., during a series of advertisements).

[0070] Depending on the implementation, the microphone may record audio samples at regular intervals (i.e., periodically) or at irregular intervals (e.g., aperiodically or with a time-varying period). For instance, the microphone may acquire audio data throughout the day at a constant rate (e.g., about two samples per minute). In other cases, the microphone may operate at one sampling rate when the TV is on or likely to be on (e.g., early evening) and at another, lower sampling rate when the TV is off or likely to be off (e.g., early morning, mid-day). If the local processor detects that the TV has been turned on (off) from the audio samples, it may increase (decrease) the sample rate accordingly. The may also trigger the image sensors to start (stop) imaging the viewing area in response to detecting that the TV has been turned on (off) from the audio samples.

[0071] As or once the raw image and audio data has been processed, the local processor overwrites the raw image and audio data or erases the raw image and audio data from memory (**230**). In other words, each image is held in the memory **150**, while the processor **150** detects and identifies humans and gauges their engagement and expressions. The detection, identification, and engagement data is collected per frame, and this information is persisted and eventually uploaded to the backend server **170**. Similarly, the audio data is also held in the memory **160**, while the third-party API is processing the audio fingerprint and returning the identity of the associated video. The identity is stored and/or uploaded to the backend server **170** as described below.

[0072] By overwriting or erasing (or otherwise discarding) the raw image and audio data, the local processor reduces demands on the memory and reduces or eliminates the ability to identify the individuals in the viewing area. This maintains the individuals' privacy by exposing less information to potential attempts to hack the system. It also eliminates the possibility of transmitting images of the individuals to third parties. This is especially beneficial for

preserving the privacy of children in the viewing area in accordance with the Children's Online Privacy Protection Act.

[0073] In some cases, the local processor actively erases the raw image and audio data from the memory. In other cases, the local processor stores that raw image and data in one or more buffers in the memory that are sized not to store more than a predetermined amount of raw image and audio data (e.g., one image or one audio sample). The local processor analyzes the raw image and data in the time period between samples so that the next image or audio sample overwrites the buffer.

[0074] The local processor **150** also stores the processed data into the memory **160**. The processed data may be stored in a relatively compact, such as comma-separated variable (CSV) format, to reduce memory requirements. The data included in the CSV or other file may indicate, for example, whether anyone is present in each image, the number of people in the viewing area **101** in each image, the number of people who are actually watching the display **11** in the viewing area **101**, the classification of each viewer's emotion, and the identity of each viewer. The processed data may also include indications about the local device's operational state, including the IR image acquisition rate, visible image acquisition rate, audio sampling rate, current software/firmware update, etc.

[0075] The local processor transmits the processed data to the remote server (e.g., via a network interface) for storage or for further processing (**236**). Because the processed data is in a relatively compact format, the upload bandwidth is much lower than it would be for raw image and audio data. And because the transmitted data does not include images of the viewing area or audio samples that could include the viewers' voices, there is less risk of compromising the viewers' privacy. In addition, the audio and image portions of the processed data are more likely to be and remain synchronized because they are processed locally than if the raw image and audio image were transmitted to and processed by a remote server.

[0076] In some cases, the local processor may transmit the processed data to the remote as it is processed. In other cases, the local processor may identify transmission windows (**234**), e.g., based on the available upstream bandwidth, the amount of data, etc. These transmission windows may be predetermined (e.g., 2 am ET), set by a household member during local device installation, set by the remote server (e.g., via a software or firmware update), or determined by the local processor based on bandwidth measurements.

[0077] FIG. **2**B illustrates a method of training a computer vision model for quantifying viewer engagement. At **241**, both the RGB and IR sensors acquire video data, which undergoes two types of processing. At **242***a*, the video data is manually annotated to identify faces in each frame. At **242***b*, a current model (e.g., a default model or a model from previous use) is sued to automatically detect faces in each frame. At **243***b*, a processor is used to compute accuracy of the automatic detection at **242***b* against the annotated videos acquired at **242***a*. At **244**, if the accuracy is acceptable, the method **240** proceeds to **245**, where the current model is set as the production model for facial recognition (e.g., used in the method **200**). If the accuracy is not acceptable, the method **200** proceeds to **243***a*, where the videos are split into a training set of videos (**246***a*) and a test set of videos (**246***b*). For example, the RGB videos can be selected as the training videos **246***a* and the IR videos can be selected as the test videos **246***b* (or vice versa).

[0078] The training videos **246***a* are sent to train a new model at **247***a*, while the test videos (**246***b*) are sent to step **247***b* for testing the new model. At **247***b*, the training videos **246***a* and the test videos **246***b* are collected together so as to compute accuracy of the new model at **247***c*. At **249**, the processor again computes the accuracy of the new model. If the accuracy is acceptable, the new model is set as the production model (**245**). If not, the method **240** proceeds to **248**, where parameters of the new model are tuned. Alternatively, another new model can be built at **248**. In any event, parameters of the new model are sent back to **247***a*, where the training videos **246***a* are used to training the new model. In this manner, a new model can be iteratively built to have an acceptable accuracy.

[0079] Remote Server Operation

[0080] In operation, the remote server **170** collects data transmitted from different local devices **105** disposed in different households. The remote server **170** can read the incoming data on a regular basis. The remote server **170** can also parse the received data and join the video recognition data with the audio recognition data using the timestamps of when each was saved.

[0081] The remote server **170** can also correct mislabeled data. For example, the remote server **170** can fix blips when a viewer is not identified or is misidentified using data from preceding and following timestamps. If a person is identified in an image preceding the image at issue and also in an image following the image at issue, the remote server **170** can determine that this person also appears in the image at issue.

[0082] The remote server **170** can also load data received from local devices **105** and/or data processed by the remote server **170** into a query-able database. In one example, the remote server **170** can also provide access to users, who can then use the stored data for analysis. In another example, the stored data in the query-able database can also facilitate further analysis performed by the remote server **170**. For example, the remote server **170** can calculate attention index and viewer index using the database.

[0083] Assessing Viewer Engagement

[0084] FIGS. **3**A-**6** illustrate methods of quantifying viewer engagement with videos using measures such as viewability index and attention index. The following definitions may be helpful in understanding the inventive methods and apparatus for quantifying viewer engagement with videos:

[0085] Program Duration is defined as the total duration of a unique program, e.g., in seconds, minutes, or hours. The actual unit (seconds, minutes, or hours) used is immaterial as long as the durations of different programs can be compared.

[0086] Commercial Duration is defined as the total duration (e.g., in seconds or minutes) of a unique commercial.

[0087] Watching Duration (Seconds) is defined as the total duration (number of seconds) that are watched of a unique program or commercial per home. Alternatively, Watching Seconds can be defined as the total duration of program in seconds minus the total time (in seconds) during which no home watches the program.

[0088] Aggregated Watching Duration (Seconds) is defined as the total duration (number of seconds) that are watched of a unique program or commercial across all homes.

[0089] Positive Duration Ratio is defined as the percentage (%) of a program or commercial advertise that has been watched. More specifically, the Positive Duration Ratio of a program or advertisement can be calculated as the ratio of the Aggregated Watching Duration over total duration of the program or advertisement times the number of households.

[0090] Viewer Count (VC) is defined as the total number of viewers in the viewing area across all homes with positive Watching Seconds for a given program or commercial advertisement.

[0091] Watching Rate (WR) is defined as the ratio of the total number of people across all homes where the TV is on over the total number of people in all households. For example, if the methods take into account 100 households having a total number of 300 people. If 30 households having 100 people have their TV set on, the watching rate is then 33.3% (i.e., 100/300). However, if the same 30 households have 150 people, then the watching rate is 50% (i.e., 150/300).

[0092] Viewing Rate (VR) is defined as the ratio of the total number of people in the viewing area across all homes over the total number of TV sets that are on. For example, if 100 people are in the viewing areas defined by 40 different TV sets (each TV set defines one viewing area), then the viewing rate is 2.5 (i.e., 100/40).

[0093] Attention Rate (AR) is defined as the ratio of the total number of people attentive to the TV across all homes over the total number of people in the viewing area across all homes. For example, if 100 people are in the viewing areas across all individuals taken into account by the methods, but only 60 people are actually watching TV (the rest 40 people may just leave the TV on while doing other things), then the attention rate is 0.6 or 60%.

[0094] Viewability Index (VI) is defined as the average of Viewing Rates (VRs) for each program and commercial.

[0095] Attention Index is defined as the average of Attention Rates (ARs) for each program and commercial.

[0096] FIG. **3**A illustrates a method **300** of assessing viewer engagement (e.g., box **206** in the method **200** of FIG. **2**A) including facial and eyeball tracking **310**, facial recognition **320**, and sentimental analysis **330**. A processor (e.g., the local processor **150** shown in FIG. **1**) can be used to implement the method **300**. The input data in method **300** can be the data acquired by the local device **105** shown in FIG. **1**, such as the image data, audio data, or depth data of the viewing area. Face and eyeball tracking **310** is employed to identify characteristic data points to track the face as it moves and determine if user is watching screen. Facial recognition **320** is employed to determine a viewer's identity using, for example, artificial intelligence. Sentimental analysis **330** is employed to determine a viewer's emotion using, for example, artificial intelligence to analyze facial features, gestures, and heart rate, among others.

[0097] The acquired information, including whether a viewer is in fact watching the screen, the identity of the viewer, and the emotion of the viewer, is used to determine various video ratings **340**. In one example, the acquired information is used to estimate individual video rating for each household. In another example, the acquired information is used to estimate individual video rating for each demographic region. In yet another example, the acquired information is used to estimate overall video rating for a group of videos. In yet another example, the acquired information is used to estimate audience reactions to specific videos (e.g., programs and advertisements). The acquired information can also be used to determine quantitative measures of viewer engagement, such as viewability index and attention index as described below.

[0098] Steps **310**, **320**, and **330** in the method **300** can be achieved using pattern recognition techniques. These techniques can determine whether any viewer is present in the viewing area by, for example, recognizing one or more human faces. If there is indeed a face recognized, these techniques can further determine who the viewer is by, for example, comparing the recognized face with a database including the facial data of the household where the video is playing. Alternatively, these techniques may use extended database to include facial data of more people (e.g., the entire community if possible) in case the viewer is not from the household. These techniques can also trace the movement of the face and analyze the orientation of the face so as to determine, for example, whether the viewer is watching the videos.

[0099] Artificial intelligence, machining learning, and trained neural network learning techniques can also be used to analyze the emotion of the viewer. To this end, these techniques analyze the body gestures (static gestures at certain time), body movements (change of gestures), facial orientations, direction/movement/positioning of faces, and heart rate, among others.

[0100] In another example, the method **300** can first recognize a face from image data acquired by, for example, the RGB sensor **140** and IR sensor **120** shown in FIG. **1**. The method **200** can also detect the position of the face, identify characteristic points on the face (e.g., boundaries points of eyes and mouth as shown in FIG. **2**A), and track the face as it moves. Using eyeball tracking techniques, the method **300** can determine whether the view is actually watching the videos (or instead just sitting in the viewing area but doing something else). Then, using techniques of trained neural network learning, the method **300** can match the viewer with a known person in the household by comparing facial features from the database in a similar position. Once the viewer has been identified, the method **300** can continually track the viewer for notable facial configurations to determine the user's mood and/or emotion.

[0101] The method **300** can also compare the audio data (e.g., acquired by the microphone **140** shown in FIG. **1**) with an audio database of videos (e.g., TV shows) and other audio so as to determinate which video is being played at a specific timing point. In one example, the video matching can determine which TV station is being viewed by the viewer(s) identified by the method **300**. In another example, the video matching can determine which TV program is being viewed by the viewer. In yet another example, the video matching can determine which commercial advertisement is being viewed. Alternatively, or additionally, the TV channel, program, or advertisement that is being viewed can be determined from data collected from other sources, including, but are not limited to, a cable or satellite set top box or other programming provider's hardware or broadcast signal.

[0102] FIG. **3**B illustrates the concepts of viewability index and attention index that can be estimated via techniques described herein to quantify viewer engagement. In

general, viewability index quantifies the propensity of what is on screen to bring people into the room. Attention index quantifies the propensity of what is on screen to engage a viewing audience. In other words, the viewability index can be regarded as the probability of a video (or other displayed content) to attract a viewer in the first place, while the attention index can be regarded as the probability of a video to keep a viewer in front of the display after the viewer is already in the viewing area. As illustrated in FIG. **3**B, the viewability index is dependent on the number of people present in the viewing area, while the attention index is dependent on the number people who are actually watching the display.

[0103] Assessing Viewer Engagement with a Viewability Index and an Attention Index

[0104] FIG. **4**A illustrates a method **401** of quantifying viewer engagement using viewability index. The method **401** can be implemented by a processor. The method **401** starts at step **411**, in which image data in acquired by the processor at each household in a plurality of households, which participate in the method via, for example, installing or using the local device **105** in the system shown in FIG. **1**. The image data includes images of a viewing area in front of a display which can play videos (e.g., TV programs, advertisement, user-request video, or any other video). In addition, the processor also determines if the display is showing a video at step **411**. At step **421**, the processor estimates the viewing rate and watching rate for each video that is played by the display. The viewing rate represents a ratio of a total number of people in the viewing areas to a total number of displays showing videos, as defined above. Similarly, the watching rate represents a ratio of total number of people in households with display showing videos to a total number of people in the plurality of households, as defined above.

[0105] The estimation of the viewing rates and the watching rates is based on the image data acquired at step **411** and on demographic information about each household in the plurality of households. The demographic information can be stored in a memory operably coupled to the processor such that the processor can readily retrieve the demographic information. In another example, the processor can acquire the demographic information from another server. At step **330**, the processor determines a viewability index based on the viewing rate and the watching rate, for each unique video in the plurality of videos. The viewability index is defined above as an average of viewing rate for each video, such as a program and a commercial.

[0106] The method **401** can further include estimating the viewer count and the positive duration ratio of each video played by the display. The estimation is based on the image data and on demographic information about each household in the plurality of households. As defined above, the viewer count represents a total number of people engaged with each unique video, and the positive duration ratio represents a ratio of total time spent by people in the plurality of households watching the unique video to a duration of the unique video.

[0107] Based on the viewer count and the position duration ratio, a balanced viewability index can be determined. In one example, the balanced viewability index can be calculated as the weighted average of viewability index (VI) by factoring in the viewer count and positive duration Ratio for each given program and commercial. In another example, the balanced viewability index can be calculated by normalizing the viewability index across the unique videos in the plurality of videos.

[0108] The method **401** can further include averaging the viewability index across all programs and commercials for a finite period of time so as to produce an average viewability index. The viewability index of each program and commercial can be divided by the average viewability index (e.g., computed on a daily, weekly, or monthly basis) so as to produce a final viewability index (dimensionless quantity) for users, such as advertising agencies, TV stations, or other content providers. In one example, the finite period of time is about two weeks. In another example, the finite period of time is about one month. In yet another example, the finite period of time is about three months.

[0109] The image data can be acquired at various acquisition rates. In one example, the image data can be taken 50 times per second (50 Hz). In one example, the image data can be taken 30 times per second (30 Hz). In yet another example, the image data can be taken every second (1 Hz). In yet another example, the image data can be taken every 2 seconds (0.5 Hz). In yet another example, the image data can be taken every 5 seconds (0.2 Hz). In addition, the method **300** can take and categorize image data for each viewer in the viewing area so as to derive viewer engagement information taking into account demographic information of the household.

[0110] FIG. **4**B illustrates a method **402** of quantifying user engagement with videos using attention index. The method **402** includes step **412**, at which image data of a viewing area in front of a display is taken for each household participating in the viewer engagement assessment. At step **412**, a processor determines whether the display is showing any video when the image data is taken (e.g., via audio data acquired by the microphone **140** in the local device **105** shown in FIG. **1**). At step **422**, for each video played by the display, the processor estimates an attention rate based on the image data and on demographic information about the household. As defined above, the attention rate represents a ratio of a total number of people engaged with the video to a total number of people in the viewing areas. Based on the attention rates of videos, an attention index is determined at step **432** to indicate the effectiveness of the video.

[0111] The method **402** can further include estimating viewer count and positive duration ratio of the video(s) played by the display. Similar to the method **401**, the method **402** can determine the viewer count and positive duration ration based on the image data and on demographic information about each household. Using the viewer count and positive duration ration, the processor can then determine a balanced attention index. The method **402** can include producing a normalized attention index by normalizing the attention index across the unique videos in the plurality of videos over a given period of time (e.g., one week, or one month).

[0112] The method **402** can further include averaging attention index across all programs and commercials for a finite period of time so as to produce an average attention index. The attention index of each program and commercial can be divided by the average attention index so as to produce a final attention index (dimensionless quantity) for customers, such as advertising agencies, TV stations, or other content providers.

[0113] Assessing Viewer Engagement Using Facial Recognition Techniques

[0114] FIG. **5** illustrates a method of assessing viewer engagement with videos using facial recognition techniques and other artificial intelligence techniques. The method **500** starts at step **510** where images of a viewing area in front of a display are captured (e.g., using the system shown in FIG. **1**). For each acquired image, the number of people in the viewing area is estimated at step **520**. In one example, the estimation can be performed using, for example, facial recognition techniques. In another example, the estimation can be performed based on body skeleton detection.

[0115] At step **530**, with respect to the display, the orientation of the face of each person in the viewing area is determined. For example, the orientation of the face can be toward the display, indicating that the viewer is actually watching the videos on the display. Alternatively, the orientation of the face can be away from the display, indicating that the viewer is not watching the video, although he or she is within the viewing area of the display. Therefore, based on the orientation of the viewers' faces, a processor can assess whether each person in the viewing area is actually engaged with the video, at step **540**. By distinguishing people actually watching the videos from those who are not watching, the processor can make more accurately determination of the effectiveness of the video. The effectiveness of the video can be quantified by, for example, how long the video can keep the viewer in an engaged state.

[0116] Detecting Skeleton, Face, Identification, Emotion, and Engagement

[0117] FIG. **6** is a flowchart illustrating a method **600** to detect skeleton, face, identification, emotion, and engagement, which in turn can be used for viewer engagement assessment described above. The method **600** can be implemented by a processor (e.g., the processor **150** or the processor in the remote server **170**). The method **600** starts at step **610**, where image data of a viewing area in front of a display is provided (e.g., by a memory or directly from the image taking device, such as the RGB sensor **130** shown in FIG. **1**). At step **620**, the processor acquires a skeleton frame (i.e., an image frame including image of at least one possible viewer, see, e.g., **230** in FIG. **2**A) from the image data. At step **630**, a processing loop is initiated, where the processor uses six individual skeleton data points/sets for each skeleton frame for further processing, including facial recognition, emotion analysis, and engagement determination. Once the skeleton data has been processed, the method **600** returns to skeleton frame acquisition at step **620** via a refreshing step **625**.

[0118] Step **635** in the method **600** is a decision step, at which the processor determines whether any skeleton is detected in the selected skeleton data in the skeleton frame. If not, the method **600** returns to step **630**, where a new skeleton data is picked up for processing. If at least one skeleton is detected, the method **600** proceeds to step **640**, where a bounding box is generated to identify head area of viewers in the image data. The bounding box can be generated based on, for example, the skeleton information, e.g., by identifying the head from the overall skeleton.

[0119] Step **645** again is a decision step, where the processor determines whether a bounding box is generated (i.e., whether a head area is detected). It is possible that an image includes an overall skeleton of a viewer but the head part of the viewer is obstructed and therefore is absent from the image. In this case, the method **600** again returns to step **630**, where the processor picked up new skeleton data. If a bounding box is detected, the method **600** goes to step **650**, where the processor carries out a second level facial recognition (also referred to as face detection). At this step, the processor attempts to detect human face within the bounding box generated at step **640**. The face detection can be performed using, for example, Haar Feature-based Cascade Classifier in OpenCV. More information can be found in U.S. Pat. No. 8,447,139 B2, which is incorporated herein by reference in its entirety.

[0120] At step **655**, the processor determines whether a face is detected at step **650**. If not, a first level facial recognition is performed at step **660**. This first level facial recognition step can be substantially similar to the second level facial recognition performed at step **650**. Performing another round of face detection may reduce the possibility of accidental failure of the facial recognition techniques. Step **665** is a decision step similar to step **655**, where the processor determines whether a face is detected.

[0121] If a face is detected at either first level facial recognition or second level facial recognition, the method **600** proceeds to step **670** to perform facial landmark detection, also referred to as facial feature detection or facial key points detection. The step **670** is employed to determine locations of different facial features (e.g. corners of the eyes, eyebrows, and the mouth, the tip of the nose, etc.). More information of facial landmark detection can be found in U.S. Patent Publication No. 2014/0050358 A1 and U.S. Pat. No. 7,751,599 B2, which are incorporated herein in their entireties.

[0122] At step **672**, the processor determines whether any facial landmark is detected at step **670**. If not, the method **600** returns to step **630** to select another skeleton data for further processing. If at least one facial landmark is detected, the processor further determines, at a decision step **674**, whether any face is detected at the second level facial recognition in step **650**. If yes, the method **600** proceeds to step **690**, where the detected face is identified (i.e., determining who the viewer is), after which the method goes to step **680**, where emotion of the face based on the facial landmark is predicted. If, at step **674**, the processor finds that no face was detected at step **650**, the method **600** directly proceeds to step **680** for the processor to estimate emotion of the viewer. Emotion analysis can be performed using, for example, a Support Vector Machine (SVM) in Open CV. More information can be found in U.S. Pat. No. 8,488,023, which is incorporated herein in its entirety.

[0123] In one example, the methods illustrated in FIGS. **3-6** analyze all available videos (including TV programs and advertisement) regardless of the duration of the video or viewer count of the video. In another example, the methods illustrated in FIGS. **3-6** perform preliminary filtering to exclude videos that are either too short or have too small a viewer count before performing the quantitative analysis of viewer engagement. In this way, the quantitative analysis can result in more statistically reliable results. For example, videos that are watched for less than a finite amount of time (e.g., less than 30 seconds, less than 20 seconds, or less than 10 seconds) can be excluded. In addition, videos that are watched by less than certain number of people (e.g., less than 20 people, less than 15 people, or less than 10 people) over a finite period (e.g., 1 month, two weeks, or one week) can also be excluded.

[0124] In one example, the methods illustrated in FIGS. **3-6** are performed over live TV programs. In another example, the methods illustrated in FIGS. **3-6** are performed over recorded TV programs. If it is recognized that the timing of a program is greater than 10 minutes shifted from its original "finger creation timestamp" (e.g., from database of TV stations), the program is determined as recorded watching. Otherwise, the program is determined as live watching.

[0125] Experimental Assessment of the Commercial Message (CM) Effect

[0126] This section describes accurate viewing data collection and analysis to examine commercial message (CM) effect management. An index termed "viewability" indicates when a person is "in front of the TV". The viewability index is created for this description and the survey that generates the data. The survey conducted for two weeks with a sample of 84 people from 30 households. CM curves are defined as patterns that show the time series curves of viewability rates between two scenes. Although the personal viewing rate of CM between scenes can be constant, the viewability rate may change. The findings show that there are 7 patterns of the CM curve. The variables of the length of CM and viewability rate can significantly contribute to the shape of the CM curve. In addition, multinomial logit model can be help in determining the CM curve.

[0127] This experiment investigated the relationship between commercial messages (CM), programs, and human viewing attitudes. The experiment also characterized the systems and methods described above. The correlation between program information, such as broadcast timing and TV stations, and viewing attitudes using statistical methods were analyzed. Currently, the personal audience rating survey used in Japan registers people through a colored button on the TV remote control and records when they press the colored button at the start and end of TV viewing. Further, the People Meter (PM) indicator records what the TV audience watched and who watched the programs (Video Research Ltd. (2014): "TV rating handbook", available at the VIDEOR.COM website in PDF format, incorporated herein by reference). However, this audience rating survey usually does not allow one to distinguish between focused and casual viewing even if the audience rating is accurately captured.

[0128] Hiraki and Ito (Hiraki, A. & Ito, K. (2000): Cognitive attitudes to television commercials based on eye tracking analysis combined with scenario, *Japanese Journal of Human Engineering*, Vol. 36, pp. 239-253, incorporated herein by reference) proposed a method for analyzing the impact of CM on image recognition using visual information based on eye movement analysis. They conducted CM viewing experiments with real CM in an environment of recreated viewing situations. According to them, auditory and visual information may interfere with commodity understanding.

[0129] In this experiment, besides personal audience ratings, an indicator of physical presence captured by the system was used to measure viewing attitudes. For example, during CM, people may leave their seats and turn their attention to one another without sitting in front of the TV. Thus, viewing attitudes during CM was statistically analyzed using two indexes-personal audience ratings and physical presence. The latter index is referred to herein as "viewability."

[0130] The viewing attitude survey experiment of 84 individuals from 30 households was conducted from mid-November to the end of November in 2014. Data was obtained 24 hours per day over 14 days.

[0131] FIG. **7** shows a schematic view of a data acquisition system **700** that measures engagement of viewers in a viewing area **701** with a program or advertisement shown on a TV **702** or other display. The system **700** includes an image sensor **710** that captures images of the viewing area **701** while the TV **702** is on. The system **700** also includes a computing device **750** that stores and processes image data from the image sensor **710** and communicates the raw and/or processed image data to a server (not shown) via a communication network.

[0132] In some cases, the computing device **750** and/or the server measures viewability in addition to personal audience ratings. Viewability indicates "being in front of the TV," and this term is defined as the audience within a distance of about 0.5 m to about 4 m from the TV with the face towards the front of the TV between 70° to the left and the right. In one example, viewability is captured at the rate of 1 second, and it denotes the number of samples for one second divided by the all the samples (84 in this case).

[0133] FIGS. **8A-8G** shows seven different shapes of CM curves, which denote the transition in the value of viewability divided by the personal audience rating. This value can indicate the percentage of people who are actually watching the TV.

[0134] To explain the differences in the shape of CM curves, classification and modeling of the data can be performed. The methods of analysis employed in this experiment are discussed below. First, the multinomial logit model (see, e.g., Agresti, A. Categorical data analysis. John Wiley & Sons (2013), incorporated herein by reference) can be employed for data modeling. Then, non-hierarchical clustering can be performed using the K-means method, at least because the sample size (1,065) is large. Next, a decision tree can be constructed. Explanatory variables are used and all samples are classified using stepwise grouping. In general, the decision tree is a classification model that expresses the plurality of classification rules in a tree structure. The Gini coefficient was used as a non-purity function.

[0135] When determining the shape of the CM curve using these methods, the analysis also considers approaches or variables that are closely related to determining the shape of the CM curve. Thus, any variables that are observed substantially simultaneously with the CM broadcast can also be included.

[0136] Data from a high viewability time range of the day is used, which, in this experiment, is six hours-from 18:00 to 24:00. The viewing attitudes towards CM from five TV stations are analyzed. The ratios of the CM curves for every TV station are shown in FIG. **9**.

[0137] In the analysis, the shape of the CM curve is the dependent variable, and it is categorized from A to G, as shown in FIGS. **8A-8G**. The explanatory variables are length of CM, television station, genre, elapsed time since the start of the program, average personal audience rating for the CM, average viewability rate of the CM, average personal audience rating for the previous scene, average viewability of the previous scene, viewability rate of the current scene divided by the personal audience rating, viewability rate of the previous scene divided by the personal audience rating, and date and day of the week. The previous scene refers to the scene between the CM and the previous CM.

[0138] The discrimination results based on the multinomial logit model are shown in TABLE 1. The discrimination rate in the multinomial logit model is 20% higher than the discrimination rate at random. The discrimination rate is particularly high when the shape of the CM curve is B or G.

[0139] In this model, seven explanatory variables are used: length of CM, TV stations, elapsed time since the start of the program, average personal audience rating for the

CM, viewability rate, viewability rate of the CM divided by the personal audience rating, and viewability rate of the previous scene divided by the personal audience rating. Of the seven variables, length of CM and TV station contribute the most to the discrimination rate.

TABLE 1

Result of the multinomial logit model

True/Prediction	A	B	C	D	E	F	G	Sum
A	34	14	13	13	48	13	1	136
B	11	114	2	44	31	15	2	219
c	14	11	21	12	17	4	4	83
D	8	57	7	86	38	7	0	203
E	17	30	10	43	110	18	0	228
F	17	42	—	17	36	37	3	152
G	0	16	1	4	7	8	8	44
Sum	101	284	54	219	287	102	18	1065
Discrimination Rate	33.66	10.14	38.89	39.27	38.33	36.27	44.44	38.50

[0140] The explained variables of the seven shapes can also be stratified. Although several different kinds of stratifications can be considered, for efficient examination, the following two kinds of stratifications were compared.

[0141] Stratification 1: Monotonic shape types (C/D/E) and non-monotonic shape types (A/B/F/G). First, monotonic shape types that do not have extreme values and non-monotonic shape types that do have extreme values were stratified. The multinomial logit model to each group is applied, and then the discrimination rate for each group can be calculated. The discrimination results of stratification 1 are shown in TABLE 2. The discrimination rate of the monotonic shape type is 59.34%, while that of the mono-tonic shape type is 51.72%, and the overall discrimination rate is 53.62%.

[0142] After stratifying the monotonic and non-monotonic shape types, the overall discrimination rate is 15% higher than that in the multinomial logit model without stratification. Compared to the multinomial logit model without stratification, the difference in the discrimination rates between the shapes of the CM curve could be determined correctly (D/E/G) and incorrectly (C).

[0143] The selected explanatory variables are as follows. In the monotonic shape types, six variables are selected, namely, TV station, elapsed time since the start of the program, average personal audience rating for the CM, viewability of the CM, viewability of the previous scene, and viewability of the previous scene divided by the personal audience rating. In the non-monotonic shape types, the six variables selected are length of CM, TV stations, elapsed time since the start of the program, average personal audience rating for the CM, viewability rate of the CM, and viewability rate of the previous scene. Length of CM, which contributes to the multinomial logit model without stratification, is not selected in the monotonic shape types.

TABLE 2

Discrimination results of stratification 1

True/Prediction	A	B	C	D	E	F	G	Sum
A	67	42	0	0	0	26	I	136
B	26	169	0	0	0	24	0	219
C	0	0	IS	25	43	0	0	83
D	0	0	10	139	54	0	0	203
E	0	0	14	63	151	0	0	228
F	30	75	0	0	0	26	4	152
G	4	22	0	0	0	14	4	44
Sum	127	308	39	227	248	90	6	1065
Discrimination Rate	52.76	54.87	38.46	61.23	60.89	28.89	66.67	53.62

[0144] Stratification 2: Simple shape types (A/B/C/D/E) and complicated shape types (F/G). Second, simple shape types can be stratified, which have at most one extreme value, and complicated shape types, which have more than one extreme value. The discrimination results of stratification 2 are shown in TABLE 3. The discrimination rate of the simple shape type is 46.50%, while that of the complicated shape type is 77.55%, and the overall discrimination rate is 52.21%.

[0145] For the simple shape types, nine variables are selected-length of CM, TV station, elapsed time since the start of the program, average personal audience rating for the CM, viewability rate of the CM, average personal audience rating of the previous scene, viewability rate divided by the personal audience rating of the CM, viewability of the previous scene divided by the average personal audience rating, and date. Further, for the complicated shape types, only one variable is selected-TV stations. As this model has only one variable, all samples are classified under F. For the simple shape types, the selected variables are similar to that of the multinomial logit model without stratification.

TABLE 3

Discrimination results of stratification 2

True/Prediction	A	B	C	D	E	F	G	Sum
A	39	19	13	20	45	0	0	136
B	15	121	2	46	35	0	0	219
C	12	15	23	12	21	0	0	83
D	11	so	7	103	32	0	0	203
E	22	38	10	40	118	0	0	228
F	0	0	0	0	0	152	0	152
G	0	0	0	0	0	44	0	44
Sum	99	243	55	221	251	196	0	1065
Discrimination Rate	39.39	49.79	41.82	46.61	47.01	77.55	0.00	52.21

[0146] Cluster analysis using the explanatory variables can be performed. The discrimination results of the cluster analysis are shown in TABLE 4. The discrimination rate is 15.77%, and there is no difference in the discrimination rate between cluster analysis and random selection. In other words, in the nonhierarchical cluster analysis, the CM curve could not be classified.

TABLE 4

Discrimination results of cluster analysis

True/Prediction	A	B	C	D	E	F	G	Sum
A	10	21	10	14	58	14	9	136
B	22	25	19	11	116	16	10	219
C	6	10	4	11	38	10	4	83
D	17	28	6	10	110	25	7	203
E	32	29	10	13	109	28	7	228
F	11	29	7	16	76	9	4	152
G	4	7	2	3	26	1	1	44
Sum	102	149	58	78	533	103	42	1065
Discrimination Rate	9.80	16.78	6.90	12.82	20.45	8.74	2.38	15.77

[0147] FIG. **10** shows a classification model through a decision tree. The determination results of the decision tree are shown in TABLE 5. The discrimination rate of the decision tree is 40%. From TABLE 5, one can see that the discrimination rate of G is 0%, but that of D is higher than that of other CM curves by as much as 73%. The discrimination rate of the decision tree is slightly higher than that of the multinomial logit model without stratification.

[0148] From FIG. **10**, the characteristics of each shape of the CM curve can be identified. Shape A occurs when the viewability rate is high. Shape B occurs when the viewability rate is low and the length of CM is long. Shape C occurs when the viewability rate of a scene is not very different from that of the previous scene. Shape D occurs when the viewability rate is low and the length of CM is short. Shape E occurs when the viewability rate of the previous scene is low and the length of CM is short. Shape F occurs when the viewability rate of a scene is low while the viewability rate of the previous scene is high.

TABLE 5

Discriminant results of the decision tree

True/Prediction	A	B	C	D	E	F	G	Sum
A	17	10	14	0	32	63	0	136
B	4	121	5	8	46	35	0	219
C	5	3	31	1	34	9	0	83
D	6	70	4	30	68	25	0	203
E	5	17	8	2	128	68	0	228
F	6	29	2	0	16	99	0	152
G	2	11	2	0	7	22	0	44
Sum	45	261	66	41	331	321	0	1065
Discrimination Rate	37.78	46.36	46.97	73.17	38.67	30.84	0.00	40

[0149] Comparison and consideration. The discrimination rate by each method is summarized in TABLE 6. The method of stratification 1 has the highest rate among all methods. However, since the explained variables were stratified, it is impossible to verify the entire connection.

TABLE 6

Summary of discrimination rates

Multiple Logit	Stratification1	Stratification2	Cluster Analysis	Decision Tree
38.5	53.62	52.21	15.77	40

[0150] The discrimination rate of the multinomial logit model without stratification is almost the same as the rate of the decision tree. Because the decision tree is determined by whether or not the viewability rate is higher than a fixed value, it is difficult to understand intuitively, and the fixed value is not replicable. Therefore, the most suitable method to determine the CM curve is the multinomial logit model without stratification.

[0151] In all the methods, the variables of length of CM and viewability rate contribute the most to determining the CM curve. Therefore, TV viewing attitudes do not depend on the genre and broadcast time of the program, but on the length of CM and the viewability rate of the current and previous scenes.

[0152] In these five methods, the variables of length of CM and viewability rate greatly contribute to determining the CM curve. In this regard, two points are considered: 1) the relationship between the length of CM and viewability rate, and 2) in what kinds of situations the viewability rate is high.

[0153] The relationship between the length of CM and viewability rate is illustrated in FIG. **11**. In general, the shorter the length of CM, the higher the viewability rate is. The longer the CM, the lower the viewability rate, because people will become uninterested and stop looking at the TV.

[0154] Further, what kinds of situations lead to a high viewability rate was investigated. When little time elapses after the program begins (depending on the genre), the viewability rate is high. As TABLE 7 shows, there are noticeable differences between the average viewability rates of each genre. The viewability rate of news programs is low, whereas that of movies and music is high. FIG. **12** shows the correlation between elapsed time since the start of the program and the viewability rate. From FIG. **12**, one can see that the viewability rate is higher when shorter time has elapsed since the start of the program.

TABLE 7

Average viewability rate by genre

Genre	Viewability
Animation/Tokusatsu	0.706
Sports	0.668
Documentary	0.907
Drama	0.807
News	0.814
Variety shows	0.988
Film	1.252
Music	1.359
Hobby/Education	0.816
Tabloid shows	0.776
All	0.939

[0155] This experimental study elucidates the relationship between CM, programs, and human viewing attitudes using an exemplary embodiment of the hardware and software components of the present invention. The most suitable method to determine the CM curve is the multinomial logit model.

[0156] The variables are analyzed that can be observed during CM to examine the relationship between the CM curve and these variables. In all the method employed, the variables of length of CM and viewability rate contribute the most to determining the CM curve. Since the discrimination rate of the monotonic shape type is high, discrimination is easier, whether unchanged or changed. In other words, the shape of the CM curve is not relevant to program characteristics such as genre and date. This indicates that when the CM broadcast time is longer, the audience gets tired of watching. Moreover, if the previous scene of the program is uninteresting to the audience, then they do not watch the next CM.

[0157] Applications of Viewer Engagement Data

[0158] FIG. **13** illustrates a system of communication of data acquired using the methods and systems described herein. The system **1300** stores and processes raw data **1310** captured from TV audience panels through the motion-sensing devices, which is transferred to the computing device **1320** such as, but without limitation, the desktop machine. Then, methods of assessing viewer engagement can be performed on, for example, desktop machines to analyze and processes the data. The methods transform the after-analyzed data into performance-based TV ratings data that can be used to determine (1) who is really watching TV (who is in the audience), (2) how often the audience members look at the TV, and (3) the audience's reaction towards the TV programs and advertisements. This processed and/or summarized data is then transferred to a central storage location **1330**, such as a server, on the cloud where third parties, including but not limited to TV advertising agencies **1340**, TV networks **1350**, and any other potential clients **1360** that might find the data useful, can conveniently access the data anytime, through the collection entity's software, an application programming interface, or a web portal, specifically developed for the collection entity's clients. Alternatively, the raw data **1310** collected by the sensors of the hardware component is transferred to a central storage **1330** on the cloud directly or indirectly through an Internet connection where it is analyzed by the software component and made available to interested third parties **1340-1360**. A third party may optionally access the raw data through the system.

[0159] FIG. **14** illustrates basic elements of an example system **1400** that can utilize the data acquired and analyzed by the systems and methods described herein. The collection entity **1430** (e.g., TVision Insights) may compensate panel members **1410** (e.g., household members) who, in exchange for compensation or volunteering, allow for the placement of the hardware components depicted in FIG. **1** to be placed

atop televisions in their household for the purpose of TV viewership data collection. Panel members may be asked to provide additional information **1420**, including but not limited to, credit card transaction data, demographic and socio-economic information, social media account logins, and data from tablets, smartphones, and other devices. This data is collected, video and IR images are recorded using the system depicted in FIG. **1**, and the video can be analyzed by the methods described in FIGS. **2-6**. Once analyzed, data describing the video may be transmitted to the collection entity **1430**, which may then sell or otherwise provide the data to advertisers **1440**, TV stations **1460**, TV agencies **1450**, and other interested third parties. Optionally, the collection entity **1430** may provide access to raw collected data for separate analysis. As part of the disclosed business model, the collection entity **1430** can motivate advertisers **1440** to encourage their TV agencies **1450** to purchase this data.

[0160] FIG. **15** illustrates big data analysis and visualization based on data acquired in methods of assessing viewer engagement. In these models **1500**, the collection entity **1520** (e.g., TVision INSIGHTS shown in FIG. **15**) can collect data from households **1510** having TV sets. In return, the participating households **1510** can receive monetary compensation (or other benefit) from collection entity **1520**. The collection entity **1520** then analyzes the data collected from the participating households using big data analysis **1530***a* and visualization techniques **1530***b* to derive information such as the effectiveness of certain TV program or advertisement. This data can be then provided to advertisers, advertising agencies, TV stations, or other content providers or promoters (collectively referred to as customers **1540**) to instruct them to improve the effectiveness of their programs. In one example, the customers **1540** can subscribe this data service to the collection entity **1520** on a monthly basis with monthly fees. In another example, the customers **1540** can buy data relating to a particular video (e.g., campaign video, special advertisement during sports events, etc.) from the collection entity **1520**.

[0161] FIG. **16** illustrates examples of collection of additional information **1600** from individuals and households (TV audiences) participating in viewer engagement data collection. The TV audiences can represent national and/or local demographics useful to interested third parties. The collection entity can collect video data **1610** and the demographic information and, packaged with data gathered by the system and analyzed by the methods regarding TV viewership, provide this information to customers for compensation. Examples of information that may be collected from TV audiences include any and all information that can be obtained through social media profiles **1620** such as, but not limited to, TWITTER, Instagram, FACEBOOK, among others. The information can further include video data and audio data **1640** obtained from the systems (including both television audio and audio such as conversation originating from individuals in the household), multi-screen data **1630** including smartphone and tablet search habits, internet search history, email account information, and credit card transaction data **1650**. This list is not exhaustive, and should not be interpreted as limiting.

[0162] The collected information and data enables a collection entity to accurately evaluate the impact of TV advertisements-including unprecedented measurements of individual demographics, which are valuable to the advertisers. The advertisers can use the data to determine which ad slots would be a best fit for their targeted audience. The message can also be more pertinent to the type of audience and can effectively lead to purchasing behavior, increasing return of investment (ROI) for the advertisers.

[0163] TV networks can also benefit from the disclosed invention as they will be able to glean more accurate ratings of their TV programs, audience type, reactions, and predictive ad slot value. This will allow them to improve their programs to better fit the type of audience and eliminate non-popular ones, in addition to determining which ad slots will have the highest value for a particular target demographic. The data can also be used to compare programs across multiple channels at the same or different time slots for a comparative evaluation of programs and advertising. Similarly, TV audience data and behavior can be collected and compared for any given programming time slot to streaming content. TV pilot programs can also be evaluated using the system before ordering episodes.

CONCLUSION

[0164] While various inventive embodiments have been described and illustrated herein, those of ordinary skill in the art will readily envision a variety of other means and/or structures for performing the function and/or obtaining the results and/or one or more of the advantages described herein, and each of such variations and/or modifications is deemed to be within the scope of the inventive embodiments described herein. More generally, those skilled in the art will readily appreciate that all parameters, dimensions, materials, and configurations described herein are meant to be exemplary and that the actual parameters, dimensions, materials, and/or configurations will depend upon the specific application or applications for which the inventive teachings is/are used. Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific inventive embodiments described herein. It is, therefore, to be understood that the foregoing embodiments are presented by way of example only and that, within the scope of the appended claims and equivalents thereto, inventive embodiments may be practiced otherwise than as specifically described and claimed. Inventive embodiments of the present disclosure are directed to each individual feature, system, article, material, kit, and/or method described herein. In addition, any combination of two or more such features, systems, articles, materials, kits, and/or methods, if such features, systems, articles, materials, kits, and/or methods are not mutually inconsistent, is included within the inventive scope of the present disclosure.

[0165] The above-described embodiments can be implemented in any of numerous ways. For example, embodiments of designing and making the technology disclosed herein may be implemented using hardware, software or a combination thereof. When implemented in software, the software code can be executed on any suitable processor or collection of processors, whether provided in a single computer or distributed among multiple computers.

[0166] Further, it should be appreciated that a computer may be embodied in any of a number of forms, such as a rack-mounted computer, a desktop computer, a laptop computer, or a tablet computer. Additionally, a computer may be embedded in a device not generally regarded as a computer but with suitable processing capabilities, including a Per-

sonal Digital Assistant (PDA), a smart phone or any other suitable portable or fixed electronic device.

[0167] Also, a computer may have one or more input and output devices. These devices can be used, among other things, to present a user interface. Examples of output devices that can be used to provide a user interface include printers or display screens for visual presentation of output and speakers or other sound generating devices for audible presentation of output. Examples of input devices that can be used for a user interface include keyboards, and pointing devices, such as mice, touch pads, and digitizing tablets. As another example, a computer may receive input information through speech recognition or in other audible format.

[0168] Such computers may be interconnected by one or more networks in any suitable form, including a local area network or a wide area network, such as an enterprise network, and intelligent network (IN) or the Internet. Such networks may be based on any suitable technology and may operate according to any suitable protocol and may include wireless networks, wired networks or fiber optic networks.

[0169] The various methods or processes outlined herein may be coded as software that is executable on one or more processors that employ any one of a variety of operating systems or platforms. Additionally, such software may be written using any of a number of suitable programming languages and/or programming or scripting tools, and also may be compiled as executable machine language code or intermediate code that is executed on a framework or virtual machine.

[0170] In this respect, various inventive concepts may be embodied as a computer readable storage medium (or multiple computer readable storage media) (e.g., a computer memory, one or more floppy discs, compact discs, optical discs, magnetic tapes, flash memories, circuit configurations in Field Programmable Gate Arrays or other semiconductor devices, or other non-transitory medium or tangible computer storage medium) encoded with one or more programs that, when executed on one or more computers or other processors, perform methods that implement the various embodiments of the invention discussed above. The computer readable medium or media can be transportable, such that the program or programs stored thereon can be loaded onto one or more different computers or other processors to implement various aspects of the present invention as discussed above.

[0171] The terms "program" or "software" are used herein in a generic sense to refer to any type of computer code or set of computer-executable instructions that can be employed to program a computer or other processor to implement various aspects of embodiments as discussed above. Additionally, it should be appreciated that according to one aspect, one or more computer programs that when executed perform methods of the present invention need not reside on a single computer or processor, but may be distributed in a modular fashion amongst a number of different computers or processors to implement various aspects of the present invention.

[0172] Computer-executable instructions may be in many forms, such as program modules, executed by one or more computers or other devices. Generally, program modules include routines, programs, objects, components, data structures, etc., that perform particular tasks or implement particular abstract data types. Typically, the functionality of the program modules may be combined or distributed as desired in various embodiments.

[0173] Also, data structures may be stored in computer-readable media in any suitable form. For simplicity of illustration, data structures may be shown to have fields that are related through location in the data structure. Such relationships may likewise be achieved by assigning storage for the fields with locations in a computer-readable medium that convey relationship between the fields. However, any suitable mechanism may be used to establish a relationship between information in fields of a data structure, including through the use of pointers, tags or other mechanisms that establish relationship between data elements.

[0174] Also, various inventive concepts may be embodied as one or more methods, of which an example has been provided. The acts performed as part of the method may be ordered in any suitable way. Accordingly, embodiments may be constructed in which acts are performed in an order different than illustrated, which may include performing some acts simultaneously, even though shown as sequential acts in illustrative embodiments.

[0175] All definitions, as defined and used herein, should be understood to control over dictionary definitions, definitions in documents incorporated by reference, and/or ordinary meanings of the defined terms.

[0176] The indefinite articles "a" and "an," as used herein in the specification and in the claims, unless clearly indicated to the contrary, should be understood to mean "at least one."

[0177] The phrase "and/or," as used herein in the specification and in the claims, should be understood to mean "either or both" of the elements so conjoined, i.e., elements that are conjunctively present in some cases and disjunctively present in other cases. Multiple elements listed with "and/or" should be construed in the same fashion, i.e., "one or more" of the elements so conjoined. Other elements may optionally be present other than the elements specifically identified by the "and/or" clause, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, a reference to "A and/or B", when used in conjunction with open-ended language such as "comprising" can refer, in one embodiment, to A only (optionally including elements other than B); in another embodiment, to B only (optionally including elements other than A); in yet another embodiment, to both A and B (optionally including other elements); etc.

[0178] As used herein in the specification and in the claims, "or" should be understood to have the same meaning as "and/or" as defined above. For example, when separating items in a list, "or" or "and/or" shall be interpreted as being inclusive, i.e., the inclusion of at least one, but also including more than one, of a number or list of elements, and, optionally, additional unlisted items. Only terms clearly indicated to the contrary, such as "only one of" or "exactly one of," or, when used in the claims, "consisting of," will refer to the inclusion of exactly one element of a number or list of elements. In general, the term "or" as used herein shall only be interpreted as indicating exclusive alternatives (i.e., "one or the other but not both") when preceded by terms of exclusivity, such as "either," "one of," "only one of," or "exactly one of" "Consisting essentially of," when used in the claims, shall have its ordinary meaning as used in the field of patent law.

[0179] As used herein in the specification and in the claims, the phrase "at least one," in reference to a list of one or more elements, should be understood to mean at least one element selected from any one or more of the elements in the list of elements, but not necessarily including at least one of each and every element specifically listed within the list of elements and not excluding any combinations of elements in the list of elements. This definition also allows that elements may optionally be present other than the elements specifically identified within the list of elements to which the phrase "at least one" refers, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, "at least one of A and B" (or, equivalently, "at least one of A or B," or, equivalently "at least one of A and/or B") can refer, in one embodiment, to at least one, optionally including more than one, A, with no B present (and optionally including elements other than B); in another embodiment, to at least one, optionally including more than one, B, with no A present (and optionally including elements other than A); in yet another embodiment, to at least one, optionally including more than one, A, and at least one, optionally including more than one, B (and optionally including other elements); etc.

[0180] In the claims, as well as in the specification above, all transitional phrases such as "comprising," "including," "carrying," "having," "containing," "involving," "holding," "composed of," and the like are to be understood to be open-ended, i.e., to mean including but not limited to. Only the transitional phrases "consisting of" and "consisting essentially of" shall be closed or semi-closed transitional phrases, respectively, as set forth in the United States Patent Office Manual of Patent Examining Procedures, Section 2111.03.

1. A method of quantifying viewer engagement with a video shown on a display, the method comprising:

acquiring, with at least one camera, images of a viewing area in front of the display while the video is being shown on the display;

acquiring, with a microphone, audio data representing a soundtrack of the video emitted by a speaker coupled to the display;

determining, with a processor operably coupled to the at least one camera and the processor, an identity of the video based at least in part on the audio data;

estimating, with the processor and based at least in part on the image data, a first number of people present in the viewing area while the video is being shown on the display and a second number of people engaged with the video in the viewing area; and

transmitting, by the processor, the identity of the video, the first number of people, and the second number of people to a remote server.

2. The method of claim **1**, wherein acquiring the images comprises acquiring a first image of the viewing area using a visible camera and acquiring a second image of the viewing area using an infrared (IR) camera.

3. The method of claim **2**, wherein estimating the first number of people in the viewing area comprises:

estimating a first raw number of people from the first image data and a second raw number of people from the second image data; and

comparing the first raw number with the second raw number to detect possible error in at least one of the first raw number or the second raw number.

4. The method of claim **1**, wherein acquiring the image data comprises acquiring images of the viewing area at a frame rate substantially equal to or greater than 20 frames per second.

5. The method of claim **1**, wherein acquiring the audio data comprises acquiring the audio data at an acquisition rate of about 0.1 Hz.

6. The method of claim **1**, wherein determining the identity of the video is based on audio signal fingerprinting.

7. The method of claim **1**, wherein estimating the first number of people present in the viewing area is based on body skeleton detection.

8. The method of claim **1**, wherein estimating the second number of people engaged with the at least on video is based on eye tracking.

9. The method of claim **1**, further comprising:

quantifying the viewer engagement of the video based at least in part on the first number of people and the second number of people at each house household in the plurality of households.

10. The method of claim **9**, wherein quantifying the viewer engagement comprises:

estimating an attention rate for the video, the attention rate representing a ratio of the second number of people engaged with the video to the first number of people in the viewing area; and

for each unique video in the plurality of videos, determining an attention index based on the attention rates of the videos in the plurality of videos.

11. The method of claim **10**, wherein the video is a unique video in a plurality of videos and the method further comprises:

estimating a viewer count and a positive duration ratio based on the image data and on demographic information about each household in the plurality of households, the viewer count representing the second number of people engaged with each unique video and the positive duration ratio representing a ratio of total time spent by people in the plurality of households watching the unique video to a duration of the unique video.

12. The method of claim **9**, further comprising:

determining an identity of each person present in the viewing area based at least in part on the image data,

wherein quantifying the viewer engagement of the video comprises quantifying the viewer engagement for each identified person.

13. The method of claim **9**, further comprising:

transmitting the first number of people and the second number of people to a remote server, wherein quantifying the viewer engagement is carried out at the remote server.

14. The method of claim **9**, further comprising:

determining whether a predetermined video in the plurality of videos is being displayed on the display based at least in part on the audio data, wherein quantifying the viewer engagement is based at least in part on whether the predetermined video is being displayed.

15. The method of claim **1**, further comprising:

storing the first number of people and the second number of people in a memory operably coupled to the processor; and

erasing and/or overwriting the image data.

16. The method of claim **1**, further comprising:

estimating an emotion of each person present in the viewing area.

17. The method of claim **1**, further comprising:

estimating demographic information for each person in the viewing area from the image data.

18. The method of claim **17**, wherein estimating the demographic information comprises estimating age, gender, ethnicity group, and facial expression.

19. A method of quantifying viewer engagement for unique videos in a plurality of videos, the method comprising:

at each household in a plurality of households, acquiring image data of a viewing area in front of a display;

determining if the display is showing a video in the plurality of videos;

for each unique video in the plurality of videos, estimating (i) a viewing rate and (ii) a watching rate based on the image data and on demographic information about each household in the plurality of households, the viewing rate representing a ratio of a total number of people in the viewing areas to a total number of displays showing videos and the watching rate representing a ratio of a total number of people in households with display showing videos to a total number of people in the plurality of households; and

for each unique video in the plurality of videos, determining a viewability index based on the viewing rate and the watching rate.

20. The method of claim **19**, further comprising:

for each unique video in the plurality of videos, estimating (iii) a viewer count and (iv) a positive duration ratio based on the image data and on demographic information about each household in the plurality of households, the viewer count representing a total number of people engaged with each unique video and the positive duration ratio representing a ratio of total time spent by people in the plurality of households watching the unique video to a duration of the unique video; and

weighting the viewability index based on the viewer count and the positive duration ratio.

21. The method of claim **20**, further comprising:

normalizing the viewability index across the unique videos in the plurality of videos.

22. The method of claim **19**, wherein acquiring the image data comprises acquiring a first image of the viewing area using an optical camera and acquiring a second image of the viewing area using an infrared (IR) camera.

23. The method of claim **19**, wherein determining if the display is showing the video is based at least in part on audio data of the viewing area via signal fingerprinting technique.

24. The method of claim **19**, further comprising:

transmitting the viewing rate and the watching rate to a remote server, wherein the viewability index is estimated by the remote server.

25. A system for quantifying viewer engagement with a video playing on a display, the system comprising:

at least one camera, disposed to image a viewing area in front of the display, to acquire image data of the viewing area;

a microphone, disposed in proximity to the display, to acquire audio data representing a soundtrack of the video emitted by a speaker coupled to the display;

a memory, operably coupled to the at least one camera and the microphone, to store processor-executable instructions; and

a processor, operably coupled to the at least one camera, the microphone, and the memory, wherein upon execution of the processor-executable instructions, the processor:

determines an identity of the video based at least in part on the audio data;

estimates, based at least in part on the image data, a first number of people present in the viewing area while the video is being shown on the display and a second number of people engaged with the video in the viewing area; and

transmits the identity of the video, the first number of people, and the second number of people to a remote server.

26. The system of claim **25**, wherein the video comprises a television program provided via a set-top box and the processor is not connected to the set-top box.

27. The system of claim **25**, wherein the at least one camera comprises a visible camera and an infrared camera and the image data comprises a first image acquired by the visible camera and a second image acquired by the infrared camera.

28. The system of claim **27**, wherein upon execution of the processor-executable instructions, the processor further:

estimates a first raw number of people from the first image and a second raw number of people from the second image; and

compares the first raw number with the second raw number to detect possible error in at least one of the first raw number or the second raw number.

29. The system of claim **25**, wherein upon execution of the processor-executable instructions, the processor:

stores the first number of people and the second number of people in the memory; and

erases and/or overwrites the image data.

30. The system of claim **25**, further comprising:

a network interface, operably coupled to the processor, to transmit the first number of people and the second number of people to a remote server.

* * * * *

EXHIBIT L

‘Attention metrics identify higher-value ad inventory, improve campaign effectiveness'

Yan Liu, CEO, TVision, explains how marketers can plan and optimize TV advertising by leveraging attention measurement in a cluttered ecosystem

by **Simran Sabherwal**
Published: Mar 18, 2021 8:40 AM | 7 MIN READ

Yan Liu, CEO, TVision, explains how marketers can plan and optimize TV advertising by leveraging attention measurement in a cluttered ecosystem. He adds that Attention Guarantee for TV is a huge step forward to bring TV ad buying more in line with the digital experience.

Excerpts:

Can you tell us about attention measurement? What is the concept and how can it benefit marketers?

TOP STORIES
LATEST FEATURED

Walk The Talk Communications expands its operations to 5 Cities
5 hours ago

Pepperfry acquires Brandmakerr, Rahul Kapuria appointed as VP
6 hours ago

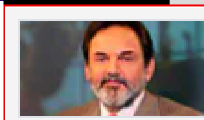
Xaxis India launches new programmatic media commerce solution
8 hours ago

Deepak Kumar Lalla takes over as MD & CEO at SBICAP Securities
9 hours ago

Spice Money elevates Kuldeep Pawar as Chief Marketing Officer
11 hours ago

SEE MORE

POPULAR
MOST READ TRENDING TOPICS


Acquisition was made without consent and without any notice: NDTV
2 days ago

Attention measurement helps marketers better understand when and where audiences are paying attention to their ads. TVision measures and reports behavioural viewing data for television (both linear and connected). This includes what people are watching, who is in the room, as well as if and when they are paying attention.

Attention measurement is especially important in today's world as people multitask and it has become harder than ever to capture their attention. Traditionally, ad buyers and sellers have bought and sold media based on ratings alone. By transacting on ratings without factoring in attention, the industry is overvaluing ads that may not be seen. By including attention metrics advertisers can identify higher-value ad inventory and improve the effectiveness of their campaigns.

Advertisers are using TVision data to build smarter, more optimized ad campaigns. They use attention data in the planning process to find highly engaged audiences that match their targeted demographics and develop day-part strategies where they'll be likely to reach attentive viewers. Marketers also use TVision data to make sure their campaign creative is effective. With TVision, they measure how well an ad captures audience attention and when the ad begins to lose its effectiveness. Because data is available immediately, advertisers can optimize these campaigns inflight.

What is the technology and how does it work?

TVision panellists set up a device near their television that is capable of picking up audio and visual signals. TVision technology identifies the specific individuals who are in the room while the TV is on. It also captures when individuals are paying attention to the TV. TVision then uses ACR to match television content with the viewing data on a second-by-second basis. By combining these data sets, the technology is able to measure how audiences pay attention to ads as well as programming across both CTV and linear TV. TVision's in-home panel is 100% opt-in and privacy-safe.

Advertisers and media sellers access the data via TVision's software as a service platform, which enables them to understand how their own content performs, as well as how competitor content performs.

How does attention-based advertising help understand viewers' behaviour and deliver an engaged audience?

After EdTech, AdTech the new favourite for investors?
3 days ago
vivo Pro Kabaddi League's Season 9 - A fiesta for brands & advertisers on TV
3 days ago
India Today Group forays into original content space; partners with Netflix & Amazon
1 week ago
Sajith Sivanandan appointed Executive VP & Head of Disney+ Hotstar
2 days ago
with Netflix & Amazon
1 week ago
Sajith Sivanandan appointed Executive VP & Head of Disney+ Hotstar
2 days ago
CATEGORY
MARKETING
with Netflix & Amazon
1 week ago
Sajith Sivanandan appointed Executive VP & Head of Disney+ Hotstar
2 days ago
CATEGORY
MARKETING
with Netflix & Amazon
1 week ago
Sajith Sivanandan appointed Executive VP & Head of Disney+ Hotstar
2 days ago
CATEGORY
MARKETING
with Netflix & Amazon
1 week ago
Sajith Sivanandan appointed Executive VP & Head of Disney+ Hotstar
2 days ago
CATEGORY

When advertisers plan their media campaigns based on Attention data they are focusing their ad spend on opportunities where they can reach engaged viewers who are actually in the room, watching the TV. Ads work when people pay attention, and as a result, these advertisers see higher levels of brand recall. There is also a much greater correlation between attention and sales, as compared to GRP and sales (see image below). Advertisers are finding that optimizing for attention ensures more effective campaigns overall.

How have clients and agencies responded to this new currency and measurement?

With TVision's solutions available in the US, UK, India, and Japan, brands and agencies around the world are embracing Attention as a new currency in their TV advertising campaigns. Major global advertisers such as Pepsi, ABInBev, and Mars have begun to include Attention data in their planning, buying and optimization processes. In Japan, TVision has been adopted by all major networks and agencies, and in the UK Confused.com and Sky are using TVision data, among other large organizations. Global agency Dentsuis one of several firms leading the charge with their focus on the Attention Economy, which is designed to bring Attention measurement to the forefront for effective cross-platform campaign measurement.

TVision is a founding member of The Attention Council which includes brands, agencies and publishers from around the world who are all focused on using attention data to build smarter campaigns.

Can you share some examples from the US market of the use case of this currency?

Recently the major beer manufacturer, Anheuser-Busch completed a

CATEGORY
MARKETING
with Netflix & Amazon
1 week ago
Sajith Sivanandan appointed Executive VP & Head of Disney+ Hotstar
2 days ago
CATEGORY
MARKETING
with Netflix & Amazon
1 week ago
Sajith Sivanandan appointed Executive VP & Head of Disney+ Hotstar
2 days ago
CATEGORY
MARKETING
with Netflix & Amazon
1 week ago
Sajith Sivanandan appointed Executive VP & Head of Disney+ Hotstar
2 days ago
CATEGORY
MARKETING
with Netflix & Amazon
1 week ago
Sajith Sivanandan appointed Executive VP & Head of Disney+ Hotstar
2 days ago
CATEGORY
MARKETING
with Netflix & Amazon

campaign with the US media company, A+E Networks. The campaign, which ran in the fourth quarter of 2020, was optimized for viewability and was the first known Attention Guarantee in the market. Anheuser-Busch ads for Budweiser ran across multiple channels, including A&E, History, FYI and Vice. Budweiser saw an increase of 7.6% in impressions served through the campaign compared to the previous quarter. You can read more about this story on our website.

In your report, you've mentioned that not all GRPs are equal and that high-attention ads need a lower number of GRPs vs low attention ads. Can you elaborate?

If you optimize your ad campaign for high attention, you need 3x fewer GRPs to reach the same level of brand awareness. For example, a campaign that reports low attention will require 2200 GRPs to reach 40% recall. But the same ad, optimized to run in high attention environments can reach 40% brand recall at just 600 GRPs. Brands that want to reduce their overall ad spend can use this approach to maintain the same level of effectiveness, or they can continue at the same ad spend and exponentially increase their effectiveness by 3x.

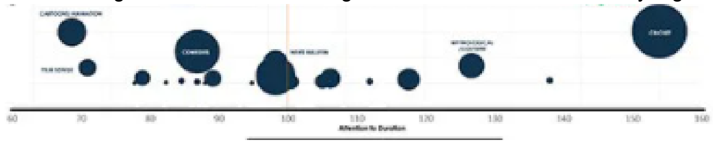
Sajith Sivanandan appointed Executive VP & Head of Disney+ Hotstar
2 days ago

CATEGORY
MARKETING
with Netflix & Amazon
1 week ago
Sajith Sivanandan appointed Executive VP & Head of Disney+ Hotstar
2 days ago

CATEGORY
MARKETING
with Netflix & Amazon
1 week ago
Sajith Sivanandan appointed Executive VP & Head of Disney+ Hotstar
2 days ago

CATEGORY
MARKETING
with Netflix & Amazon
1 week ago
Sajith Sivanandan appointed Executive VP & Head of Disney+ Hotstar
2 days ago

CATEGORY
MARKETING
with Netflix & Amazon
1 week ago
Sajith Sivanandan appointed Executive VP & Head of Disney+ Hotstar



TVision has run a pilot concept with TAM Media, here in India as well. Please do share the key insights and trends from this pilot.

TVision ran a pilot program in India to illustrate the value of Attention in TV advertising. The pilot study uses data from our panel in India. The report examines how Attention is impacted by a variety of lenses, such as daypart, genre, as well as pod duration and can be used by advertisers to optimize campaigns for improved brand recall and sales outcomes.

Advertisers will find some key takeaways in the report such as prime, and late-night dayparts deliver better than average attention across all genres; shorter pod lengths deliver better attention, and shorter ads are more efficient at capturing attention, while longer ads do deliver more total seconds of attention overall. Cricket is a winner across the board as its combination of high attention and high GRPs make it valuable programming for advertisers looking to reach large, engaged audiences.

The full report can be downloaded from the TVision website.

What are the key factors of attention that media buyers should keep in mind?

The biggest thing advertisers should keep in mind when thinking about Attention is just how much it varies. Attention varies within daypart, across networks, even across shows and ad pods. Brands need to identify their audience and find the right opportunities to reach that specific audience when the audience is highly engaged. Many factors influence attention to advertisements, including the strength of the creative and whether it contextually matches the programming content.

Brands that use Attention data to inform their campaigns layer the information on top of GRPs in order to find high quality shows- those shows that deliver attentive viewers, but also reach their target audience.

You've compared data from the US super bowl and IPL India. What are your observations about the same?

CATEGORY
MARKETING

with Netflix & Amazon
1 week ago

Sajith Sivanandan appointed Executive VP & Head of Disney+ Hotstar
2 days ago

CATEGORY
MARKETING

with Netflix & Amazon
1 week ago

Sajith Sivanandan appointed Executive VP & Head of Disney+ Hotstar
2 days ago

CATEGORY
MARKETING

with Netflix & Amazon
1 week ago

Sajith Sivanandan appointed Executive VP & Head of Disney+ Hotstar
2 days ago

CATEGORY
MARKETING

with Netflix & Amazon
1 week ago

Sajith Sivanandan appointed Executive VP & Head of Disney+ Hotstar
2 days ago

CATEGORY

Super Bowl 2021 and the IPL both delivered attention that exceeds industry norms. Ads that aired within the Super Bowl had 61% higher attention than industry norms, while ads that aired throughout the IPL had 41% higher attention than the industry.

The Super Bowl is America's biggest advertising night - but that's the thing, it is just one night. Brands pay upwards of $5 million in order to advertise in the Super Bowl, and they want to make sure it was worth it.

The unique aspect about as we see it is that IPL is just as big as the Super Bowl, but its not just one-day attention that it delivers; it delivers high attention spread out over many days, providing more opportunities for brands to advertise.

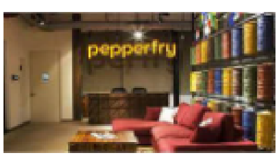
Read more news about (marketing news, latest marketing news,internet marketing, marketing India, digital marketing India, media marketing India, advertising news)

For more updates, be socially connected with us on Instagram, LinkedIn, Twitter, Facebook & Youtube

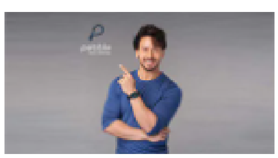
TAGS Ipl | Super Bowl | Yan Liu | TVision | Attention Guarantee | Attention Metrics | Internet Advertising | Internet Advertising India | Marketing | Internet Marketing | Marketing Campaign | Marketing Ad | Internet | Internet Ad | Internet News | Marketing Advertising India | Marketing India

Show comments

YOU MAY ALSO LIKE



Pepperfry acquires Brandmakerr, Rahul Kapuria appointed as VP
6 hours ago



Wearable brand Pebble ropes in Tiger Shroff as brand ambassador
12 hours ago

Lancer Capital appoints Rise Worldwide to manage commercial rights of UAE T20 team
13 hours ago

MARKETING

with Netflix & Amazon
1 week ago

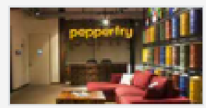
Sajith Sivanandan appointed Executive VP & Head of Disney+ Hotstar
2 days ago

CATEGORY
MARKETING

1 week ago

Sajith Sivanandan appointed Executive VP & Head of Disney+ Hotstar
2 days ago

CATEGORY
MARKETING



MARKETING
Pepperfry acquires Brandmakerr, Rahul Kapuria appointed as VP
6 hours ago

MARKETING
Wearable brand Pebble ropes in Tiger Shroff as brand ambassador
12 hours ago

MARKETING
Lancer Capital appoints Rise Worldwide to manage commercial rights of UAE T20 team
13 hours ago

MARKETING
TheSmallBigIdea bags integrated digital marketing mandate for Sandu Pharmaceuticals
14 hours ago

MARKETING
HI LIFE grants Connect Digital digital marketing AOR
14 hours ago

EXHIBIT M

TVISION SOLUTIONS INSIGHTS SUCCESS STORIES COMPANY LOGIN CONTACT US

SEE HOW PEOPLE REALLY WATCH TV

Access actionable insights and detailed data on viewer engagement across devices, apps, networks, shows, and ads

GET STARTED

ATTENTION PLATFORM

Know When TV Audiences are Paying Attention

We're the only company capable of measuring real-world viewability and attention for TV. Marketers, networks, apps, and industry watchers all use TVision's SaaS platform to find how TV content and ads engage audiences.

LEARN MORE

PANEL DATA

Strengthen and Expand Your Data

The next generation of TV currencies and measurement platforms are being built with data from TVision's panel. TVision data isn't a black box – we're proud to support a more transparent future for TV measurement.

LEARN MORE

TVision is Trusted By

VIEW ALL

ABInBev

"TVision is now an integral part of our measurement ecosystem."

Paolo Provinciali
Head of US Media
Anheuser-Busch

"Attention is informing buyin decisions at brands and agen it's powerful for our sales te this data to share."

JT Peace
Associate Director of Ad Sales

INNOVATIVE TECH + PROPRIETARY PANEL POWER OUR NEXT-GEN SOLUTIONS

TVision gathers second-by-second data from a nationally representative panel of households who have signed on to help our industry understand how, what, and when they watch TV. We turn privacy-safe observations into actionable media insights through a combination of sophisticated in-home computer-vision technology, IOT, and advanced data processing.

EVERY SECOND, WE'RE CAPTURING AND REPORTING:

What program or ad is playing on the TV

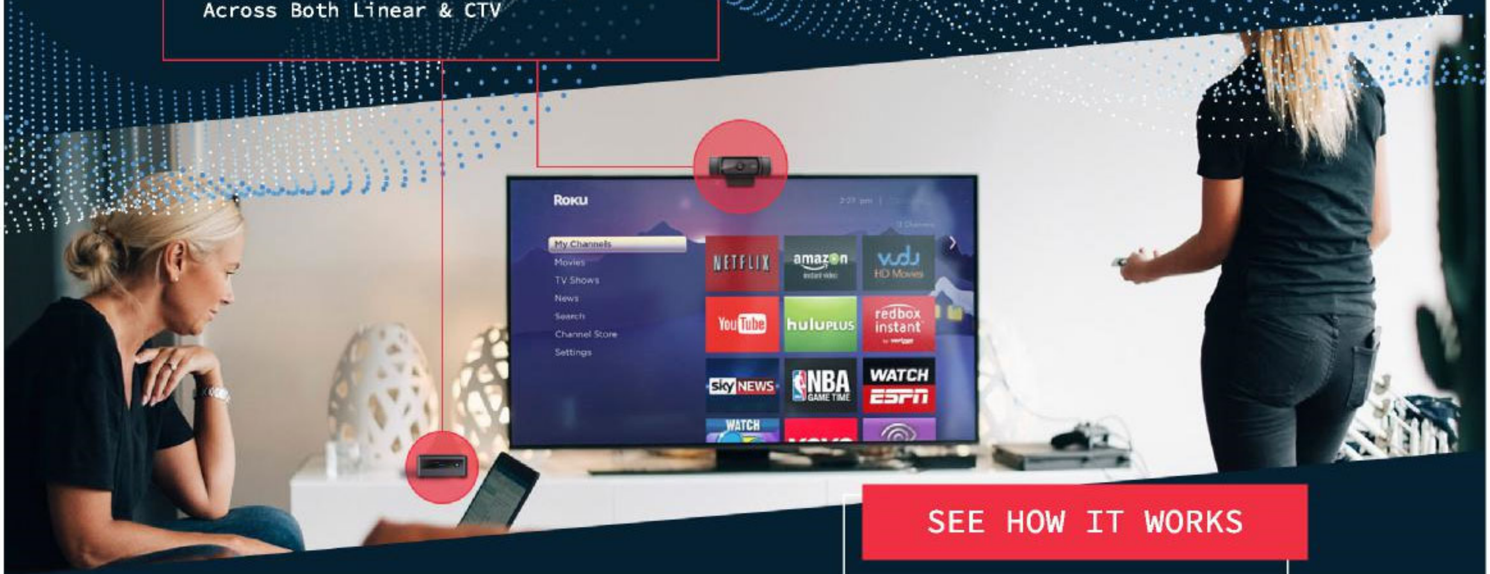
How that content is getting to the TV

Which individuals are in the room

If they're paying attention to the TV

24,783,972,198

Measured Seconds of TV Viewing Across Both Linear & CTV



SEE HOW IT WORKS

Our Solutions

FOR BRANDS & AGENCIES

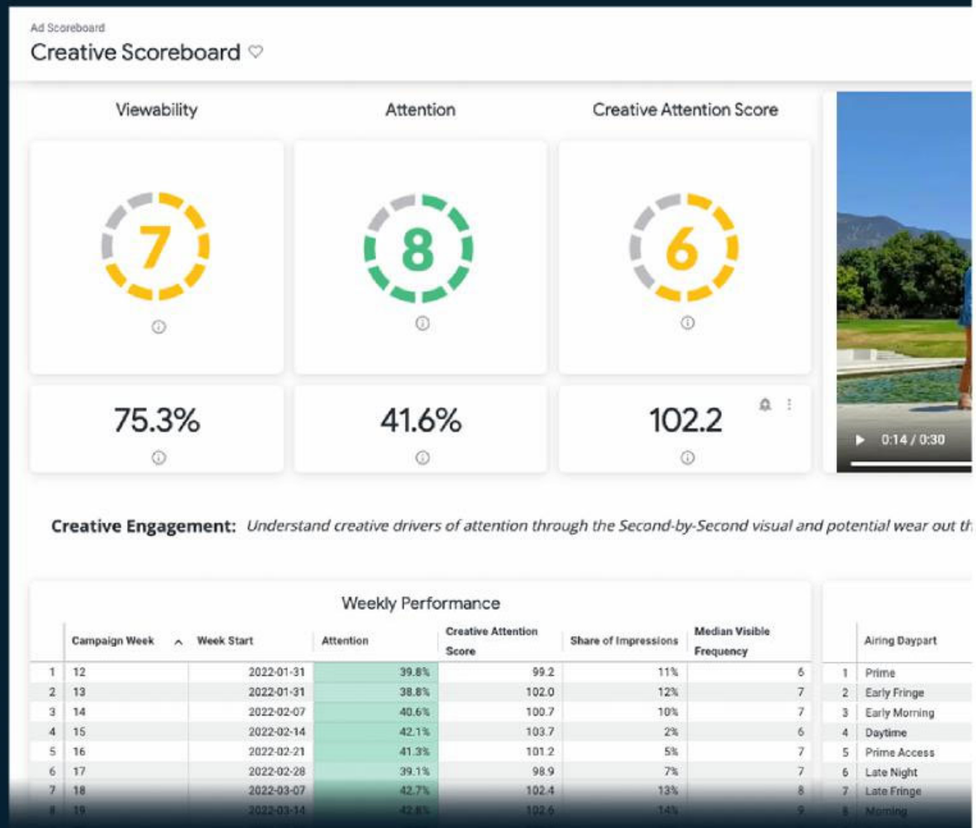
- AD SCOREBOARD
- TV PLANNING & MEASUREMENT
- DIGITAL ACTIVATION
- CREATIVE MONITORING

FOR NETWORKS & STREAMING SERVICES

- LINEAR TV INSIGHTS
- CTV ANALYTICS

FOR MEASUREMENT & DATA PARTNERS

- CALIBRATION DATA
- DATA LICENSING



CLIENT CASE STUDIES

EMAIL ADDRESS:*

SUBSCRIBE TO NEWSLETTER

EXHIBIT N

TVISION

TVision Methodology Overview

By TVision | March 10

TVision measures and reports on television viewability and attention in the USA and several international markets. Behavioral viewing data - what people are watching, who is in the room, as well as if and when they are paying attention - is collected from an opt-in, privacy safe panel. TVision uses cutting edge technology to measure person-level movements, and interprets the audio from the television to determine what content is being shown.

TVision uses three KPIs to monitor its data quality:

InTab

To monitor the health of the reporting measurements

In-balance boundaries

To monitor the health of the reporting sample

In-stability boundaries

To monitor the health of the reporting panel (the sampling over time)

ABOUT OUR PANEL

TVision recruits households from across the country. Six household characteristics and five individual characteristics are used to ensure panelists in the Northeast, South, Midwest, and West proportionately represent their market areas. These characteristics are used to project television activity and viewing behavior for the total US population. TVision uses the US Census and annual updates from the American Community Survey (ACS) to define population universe estimates (UEs) for each characteristic. In order to properly project television activity and viewing, TVision applies a daily weighting methodology, so both single-day and multi-day projections remain consistent and representative.

MEASUREMENT

In panel homes, TVision monitors television sets and individuals near the television sets second-by-second to determine:

- **Presence** - the amount of time that a viewer is in the room during a piece of content.
- **Attention** - The amount of time that a viewer is looking at the TV screen during a piece of content.

Panelists set up the device near the TV to pick up its audio signals. The microphone array on the device collects audio fingerprints that help to identify the content that is playing through the TV. These fingerprints are compared with program and ad databases, and logged as such. This data is centrally maintained in TVision's event data storage.

Individual viewing behavior is monitored with optical sensors in panel homes. The sensor signals are processed to identify the specific individuals who are in the room while the TV is on. It also captures when individuals are looking at the TV (attention). This viewing and attention data is centrally maintained in TVision's event data storage. Viewing data is matched to content data to obtain the final metrics. No video streams are stored on the device or transmitted back to TVision.

INTAB

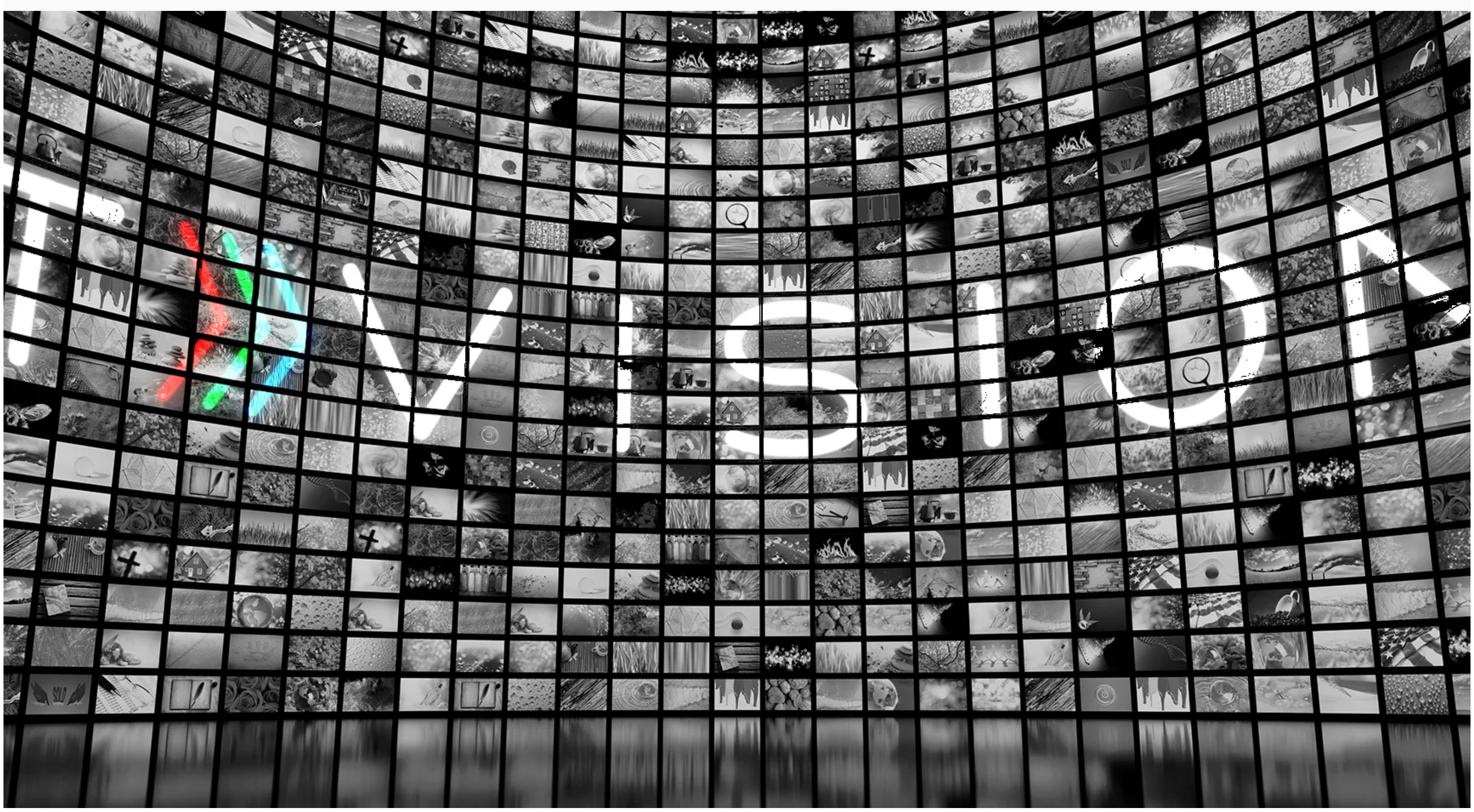
TVision evaluates the quality of data from every panel household every day. Households where the monitoring equipment and software are functioning properly within defined fault tolerances are qualified and included in the daily InTab. Households that fail to meet fault tolerance thresholds for specific days are not included in that day's InTab.

PANEL PRIVACY & SAFETY

TVision's in-home panel is 100% opt-in and privacy-safe. TVision policy prohibits any audio and images from leaving the home once a household is InTab. If equipment is removed from a home for any reason, any personal data is deleted.

For more detailed information about TVision's methodology, please contact us.

More resources from TVision



BLOG INSIGHTS

Catch Up With TVision In-Person this Fall

We are excited to get together in-person with industry leaders to discuss the future of TV and CTV and advertising attention. Catch us at TV of Tomorrow, Advertising Week, Brand Innovators, ANA Masters of Marketing, and ARF's OTT 2022.

READ MORE »

BLOG INSIGHTS

Pre Mid-Terms: Political Ads Capture Attention

As the mid-term elections approach, candidates and Super Pacs will spend an estimated $2.1 billion on linear TV and approximately $300 million on CTV advertising. Here we take a look at how effective political TV and CTV advertising is at capturing attention of the Democrat and Republican base, and engaging those crucial independent voters.

READ MORE »»

INSIGHTS

The Top Ads for Attention June - August 2022

Disney+, Geico, Apple, and YouTube TV broke through to capture TV audiences' attention best between June and August, 2022. See how the ads perform with key demos and identify the creative elements that work best with this second-by-second ad analysis.

READ MORE »»

INDUSTRY REPORT

The Insurance TV Attention Report

How do Insurance Brands Capture Attention?

DOWNLOAD NOW »»

Get in Touch

- [x] YES, SEND ME TVISION NEWS AND CONTENT.
- [ ] BRAND & AGENCY SOLUTIONS

- [ ] NETWORK & STREAMING SERVICE SOLUTIONS
- [ ] MEASUREMENT & DATA SOLUTIONS
- [ ] TVISION PANEL SUPPORT
- [ ] MEDIA INQUIRIES
- [ ] JOBS AT TVISION
- [ ] OTHER

Stay Up to Date

The TV landscape is changing quickly. Let us help keep you in the know.

FULL NAME:

EMAIL ADDRESS:*

SUBSCRIBE TO NEWSLETTER

PRIVACY POLICY
TVISION JAPAN
TVISION INDIA

EXHIBIT O

TVISION

ADVANCING TV MEASUREMENT WITH POWERFUL TECH

TVision's innovative technology is pioneering our industry's path to smarter and more accurate measurement

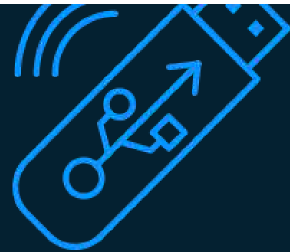
GET A DEMO

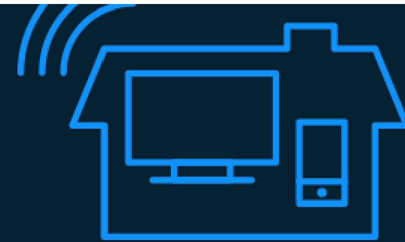
Creative Engineering Solves Complex Problems

TVision’s end product is a SasS platform, but we’re not your typical SaaS company. Our SaaS platform is powered by proprietary hardware, software, and advanced data models. We’ve developed a privacy-safe combination of sophisticated in-home computer-vision technology, IOT, and AI-powered data processing to measure TV viewing

It Starts With Understanding How People Really Watch TV

TVision’s in-home hardware is independently installed by our panelists. It is a privacy-first solution that provides unprecedented insight into how people really watch TV.





TVision Sensor

Our Sensor uses sophisticated facial recognition to detect who is in the room and matches data with ACR technology to understand what is on the TV.

TVision Digital Meter

Our Digital Meter uses network traffic analysis to detect if a digital device is connected and identifies if any streaming apps are in use.

TVision Measurement Engine

Our measurement engine remotely manages and supports thousands of in-home devices. It also identifies the source of content at any given second.

We Make Sense of a Challenging Data Set

Every second, we gather data on how each of our panelists watch TV. TVision's in-house developed and trained models help us understand who is watching, and if they are paying attention. We match and classify what they're watching with an extensive library of programming and ads. ACR and local network analysis ensures we can identify how and where that content is being delivered to the TV.

130 National channels and major local broadcasters

1100 Different streaming device models

28K Streaming titles

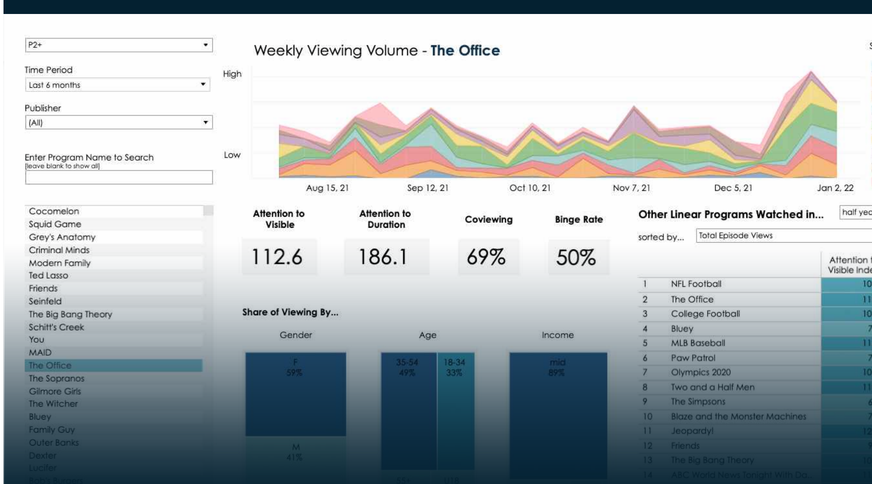
6M Program impressions across linear and CTV

69M Ad Impressions across linear and CTV

24,982,982,928
Measured Panel Seconds

We Turn Second-by-Second Viewing Data Into Powerf

The end result is our industry's only single-source platform for understanding how people really watch both linear and CTV. Our SaaS reporting allows marketers and media sellers to leverage the powerful data our technology collects.



CTV ANALYTICS

TVision's CTV Analytics dashboard breaks down CTV's walled gardens and enables media sellers and market researchers to identify the CTV devices, apps and programming that are driving adoption and engagement. A user-accessible interface delivers powerful insights on co-viewing, attention, share of time spent, and more.

LEARN MORE

LINEAR TV INSIGHTS

TV PLANNING & MEASUREMENT

AD SCOREBOARD

Innovative Technology, Inspired Engineers

TVision's technology-based approach to second-by-second TV measurement is changing the decades-old way that the TV industry buys, sells and values its inventory. Leveraging the latest engineering best practices and sophisticated technical solutions, including AI and computer vision, our engineering team is inspired everyday to design hardware and software that provides an elegant solution to a complex problem.

TVision's story is one of disruptive innovation. We're a fast-paced, lean organization with large and clear goals.

Devon Bray
Senior Software Engineer

It's challenging work. Every day I learn something new. The dream of every developer is to investigate and use the latest technology. This is what you do at TVision.

Sebastian Davalle
Senior Software Engineer

We're nimble, supportive, and innovative.

Kyle McCl ke
Head of Data Engineering

JOIN THE ENGINEERING TEAM

Stay Up to Date

The TV landscape is changing quickly. Let us help keep you in the know.

FULL NAME:

EMAIL ADDRESS:*

SUBSCRIBE TO NEWSLETTER

Privacy Policy | TVision Japan | TVision India

EXHIBIT P

TVISION

The TVision Panel enables us to measure "eyes on screen" attention to every second of programming and advertising on television.

In order to build our panel, we recruit households like yours to install TVision technology in their homes. This system uses proprietary computer vision technology to measure attention to TV shows and commercials.

FAQ

+ Does streaming count as TV?

The TVision System is not limited to broadcast TV. If you stream content instead of watching a TV channel, no problem - we can detect that. We can also detect content you save on your DVR and watch later.

+ Where does the system get set up?

The TVision system should be set up on your primary TV, the one that is watched most often. As a standalone technology, it will not interfere with your TV in any way. The system's webcam is placed above or below your TV with a mount clip.

+ How does the system get set up?

Once you've signed the online participation agreement, we will ship you a kit from our warehouse. You connect the TVision System to your home Internet network using an Ethernet cable or you connect it to your home's WiFi network using a phone or laptop to tell it the WiFi password.

+ What will the TVision system detect?

TVision technology measures if you're in the room and looking at the screen. To do this, the system starts out in Training Mode, where it captures headshot images of your household members' faces from forehead-to-chin, ear-to-ear. It's not able to capture backgrounds or body shots. The system uses those images to tell the viewers apart and map in their demographics from your household profile. An anonymized ID is created for each person. This attention data is transmitted back to TVision in a text file. The TVision System does all the work, so rest assured, no one is ever watching you!

The system captures audio of the program airing on your TV and maps it to the audio tags associated with such programs from our database. This enables us to understand what shows and commercials are on your screen. Our system is not listening for or recording your conversations, it's only able to recognize audio tags that we have subscribed to in our database.

The system will determine if there are any streaming devices running on your network and will ask them what content they're currently playing.

All we collect is the above data – anonymous and privacy-safe.

+ How do you know what's on my television?

Using technology similar to 'Shazam,' our identifier detects TV audio by searching for small digital or audio tags that are unique to each program or ad. We then match those tags to shows and commercials in our database, before sending back the matching program or commercial name to TVision.

+ How do you keep my data secure?

TVision takes security very seriously and complies with the Children's Online Privacy Protection Act (COPPA). All viewership data is de-identified and transmitted into binary digits, and then stored in an ISO 27001:2013 Accredited AWS Database.

Personally Identifiable Information (PII) is encrypted and stored in a separate database from viewership data, for contact and shipping purposes only.

What all this technical stuff means is that we take your privacy very seriously. We abide by all government regulations and invest in the best security partners to protect the data we collect.

+ How do you use the data?

Before we do anything, your data is anonymized and aggregated into averages and indexes. Then, we analyze the data to help TV networks and advertisers understand how people are watching TV content, empowering our clients to improve their programming for everyone who watches it. Here are some examples of the kinds of analysis we create from this data:

Top TV Shows By Attention

Super Bowl LII chart

Attention Madness 2018

+ How will I be rewarded?

Monthly incentives will be sent via email in the form of a digital gift card reward link. All you have to do is watch TV and check your inbox.

+ How can I apply?

Applying to join the panel is easy. All you have to do is fill out our screener to get started!

+ What happens if I move?

Please contact cs@tvisioninsights.com so that we can help you set your system up in your new home and keep our records accurate.

+ What if I have a question? Who can I reach out to?

The TVision Panel Support Team is here to help you throughout your participation in the panel. If you have any questions, just contact us. We also monitor compliance, and we'll reach out to you to help resolve any issues. For additional questions, please contact Panel Support at cs@tvisioninsights.com.

PRIVACY POLICY ABOUT CONTACT

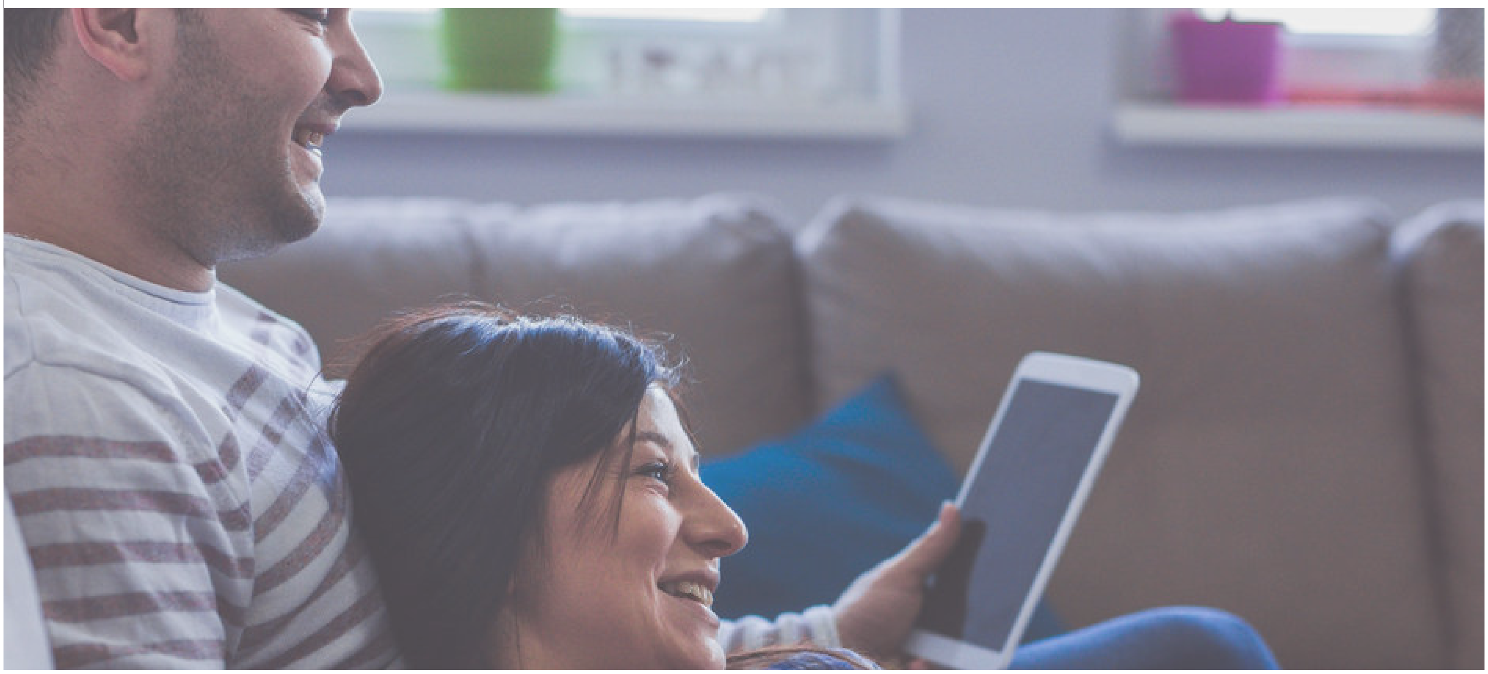


EXHIBIT Q

<u>Claim Charts for U.S. Patent No. 11,470,243</u>

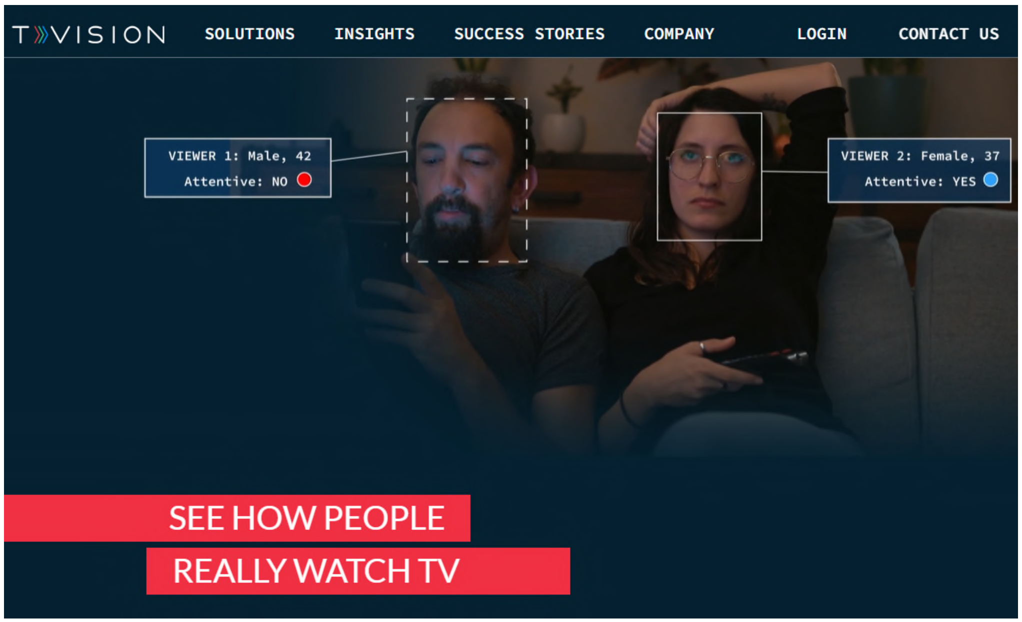
<table>
<tr><th>Element of Claim 1</th><th>TVision Insights, Inc.</th></tr>
<tr><td>An audience measurement system to obtain exposure data for a media exposure environment, the audience measurement system comprising:

memory;

machine readable instructions; and

processor circuitry to execute the machine readable instructions to:</td><td>

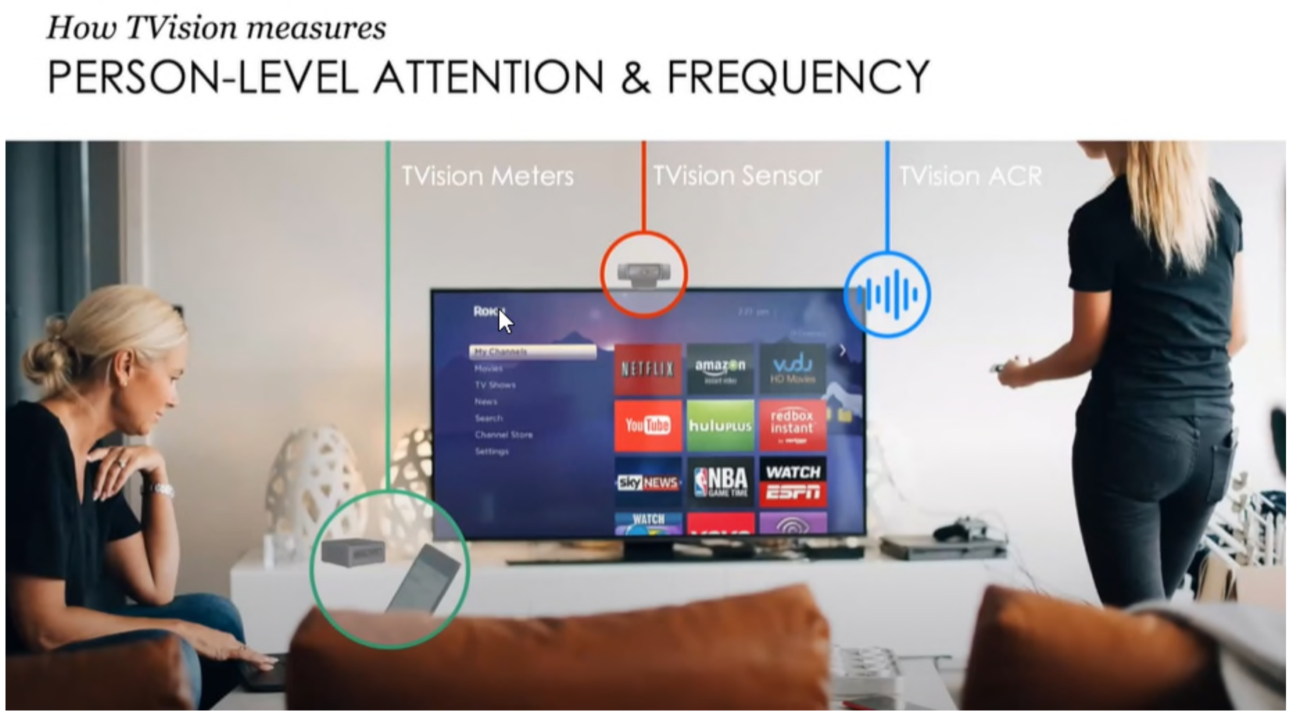
TVision "gathers second-by-second data from a nationally representative panel of households who have signed on to help our industry understand how, what, and when they watch TV." TVision captures and reports "[w]hat program or ad is playing on the TV," "[w]hich individuals are in in the room," and "[i]f they're paying attention to the TV." (*TVision, TVision Insights*, https://www.tvisioninsights.com/, Complaint Ex. M (screenshot below).)



TVision's system includes a webcam that is set up on a user's TV (*e.g.*, placed above or below the TV with a mount clip) as well as a computer that captures audio of the program or commercial airing on the TV. (*See TVision, TVision Methodology Overview*, https://www.tvisioninsights.com/resources/tvision-methodology-overview, Complaint Ex. N; *Mytvpanel, About Us*, https://www.mytvpanel.com/about, Complaint Ex. P; *TVision, Our Technology*, https://www.tvisioninsights.com/our-technology, Complaint Ex. O; "Join the TVision Panel," https://vimeo.com/295447727/3506f24b2b?embedded=true&source=video_title&owner=90679125 at 1:09 (first screenshot below); LDV Capital, "Inderbir Sidhu, CTO of TVision: Next Generation Audience Measurement,"

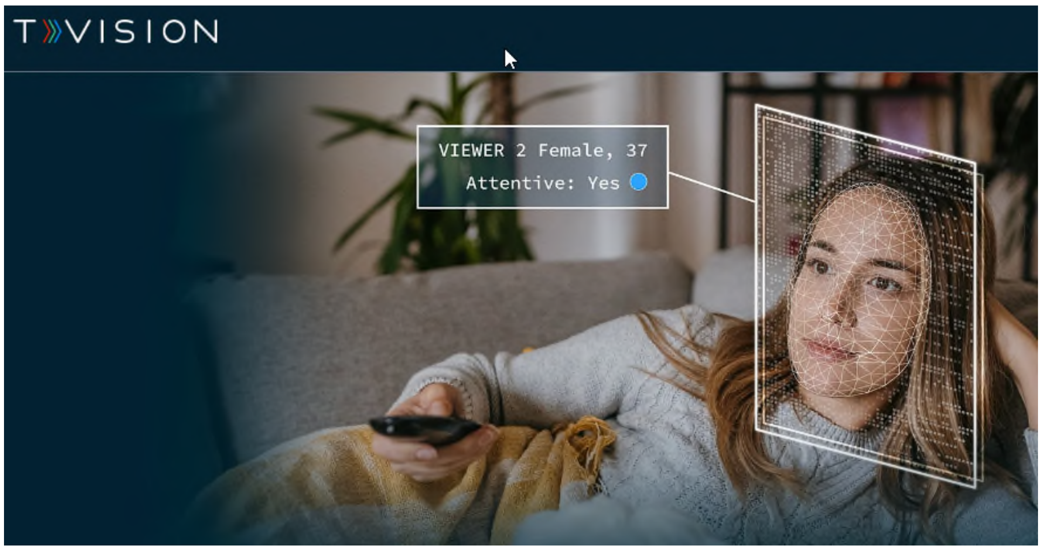
</td></tr>
</table>

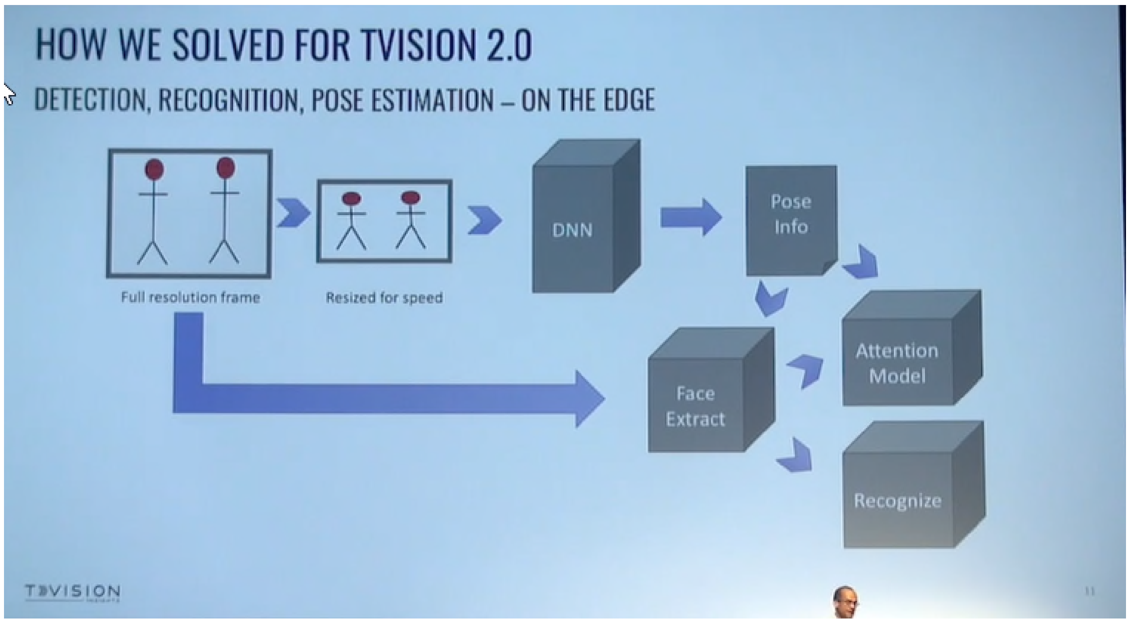
Element of Claim 1	TVision Insights, Inc.
	September 15, 2021 ("Sidhu Interview,") https://www.youtube.com/watch?v=gTBEpZo1HcM at 5:06 (second screenshot below).) 

Element of Claim 1	TVision Insights, Inc.
	 The TVision computer includes a microphone array that collects audio signals output by the TV. The TVision computer includes a processor and a memory. The memory stores instructions executable by the processor to "identify the content that is playing through the TV" using the audio signals and to monitor "individual viewing behavior" using images captured by the webcam. (*See TVision, TVision Methodology Overview*, https://www.tvisioninsights.com/resources/tvision-methodology-overview, Complaint Ex. N.)

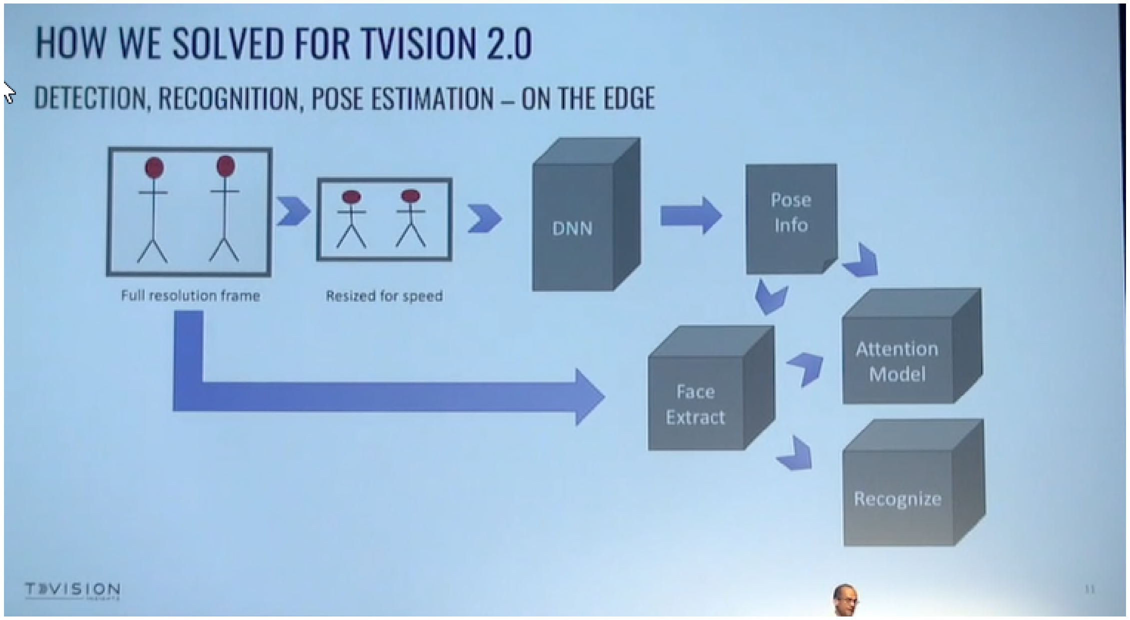
Element of Claim 1	TVision Insights, Inc.
generate an audio signature of media content presented by a television within the media exposure environment;	The TVision computer collects audio signals output by the TV and generates audio signatures (also known as fingerprints) to identify what program or commercial is on the TV. (*See TVision, TVision Methodology Overview*, https://www.tvisioninsights.com/resources/tvision-methodology-overview, Complaint Ex. N ("The microphone array on the device collects audio fingerprints that help to identify the content that is playing through the TV.")); *Mytvpanel, About Us*, https://www.mytvpanel.com/about, Complaint Ex. P ("Using technology similar to 'Shazam,' our identifier detects TV audio by searching for small digital or audio tags that are unique to each program or ad. We then match those tags to shows and commercials in our database, before sending back the matching program or commercial name to TVision."); TVision, "Join the TVision Panel," https://vimeo.com/295447727/3506f24b2b?embedded=true&source=video_title&owner=90679125 at 1:48-2:00 ("Every TV program and advertisement has a unique audio fingerprint. The TVision technology scans for these unique audio fingerprints, kind of like Shazam, and is able to identify what is on the TV.").) Shazam works by generating "a digital fingerprint" of audio captured by a device. (*See Shazam, Company*, https://www.shazam.com/company, Complaint Ex. R.) To generate audio fingerprints, the TVision computer uses software provided by ACRCloud Limited ("ACRCloud"). (*See ACRCloud, Live Channel Detection*, https://www.acrcloud.com/live-channel-detection/, Complaint Ex. S; *ACRCloud, Advertising Big Data*, https://www.acrcloud.com/advertising-big-data/, Complaint Ex. T.)

Element of Claim 1	TVision Insights, Inc.
obtain content identifying data corresponding to the presented media content, the content identifying data based on the audio signature of the media content presented by the television within the media exposure environment;	The TVision computer collects audio signals output by the TV and generates audio signatures (also known as fingerprints) to identify what program or commercial is on the TV. (*See TVision, TVision Methodology Overview*, https://www.tvisioninsights.com/resources/tvision-methodology-overview, Complaint Ex. N ("The microphone array on the device collects audio fingerprints that help to identify the content that is playing through the TV."); *Mytvpanel, About Us*, https://www.mytvpanel.com/about, Complaint Ex. P ("Using technology similar to 'Shazam,' our identifier detects TV audio by searching for small digital or audio tags that are unique to each program or ad. We then match those tags to shows and commercials in our database, before sending back the matching program or commercial name to TVision."); TVision, "Join the TVision Panel," https://vimeo.com/295447727/3506f24b2b?embedded=true&source=video_title&owner=90679125 at 1:48-2:00 ("Every TV program and advertisement has a unique audio fingerprint. The TVision technology scans for these unique audio fingerprints, kind of like Shazam, and is able to identify what is on the TV.").) Shazam works by generating "a digital fingerprint" of audio captured by a device. (*See Shazam, Company*, https://www.shazam.com/company, Complaint Ex. R.) To generate audio fingerprints, the TVision computer uses software provided by ACRCloud Limited ("ACRCloud"). (*See ACRCloud, Live Channel Detection*, https://www.acrcloud.com/live-channel-detection/, Complaint Ex. S; *ACRCloud, Advertising Big Data*, https://www.acrcloud.com/advertising-big-data/, Complaint Ex. T.) The TVision computer uses the above-described audio signatures to identify "the matching program or commercial name." (*See Mytvpanel, About Us*, https://www.mytvpanel.com/about, Complaint Ex. P.)

Element of Claim 1	**TVision Insights, Inc.**
analyze a sequence of images of the media exposure environment to detect a head appearing in one or more of the images, the sequence of images obtained by a camera while the media content corresponding to the content identifying data is presented by the television;	As explained above, the TVision computer monitors individual viewing behavior using images captured by the webcam mounted to the TV. The TVision computer then uses the images captured by the webcam to detect a user's head appearing in one or more of the images. (*See TVision, Our Technology*, https://www.tvisioninsights.com/our-technology, Complaint Ex. O (screenshot below); Sidhu Interview, https://www.youtube.com/watch?v=xnFypL2JXPE at 3:10-3:20, 4:47-5:06, and 5:32-5:47 ("It is key to us to understand the pose information . . . of the audience. . . . We use deep neural nets to get pose information. We detect humans. We track humans. We kind of have trained our own model for tracking. We use something called graph cut algorithms to track the path of people across multiple frames.").) 

Element of Claim 1	TVision Insights, Inc.
determine an orientation of the head with respect to the camera; and	The TVision computer uses computer-vision technology to determine a head pose of the detected head. (*See* Sidhu Interview, https://www.youtube.com/watch?v=xnFypL2JXPE at 5:32-5:47 ("But eventually, all of the pose data makes its way to our attention measurement. And that's where we extract out the head pose, the facial features, and all that to compute whether or not somebody is paying attention and what their gaze is.") (screenshot below).) 

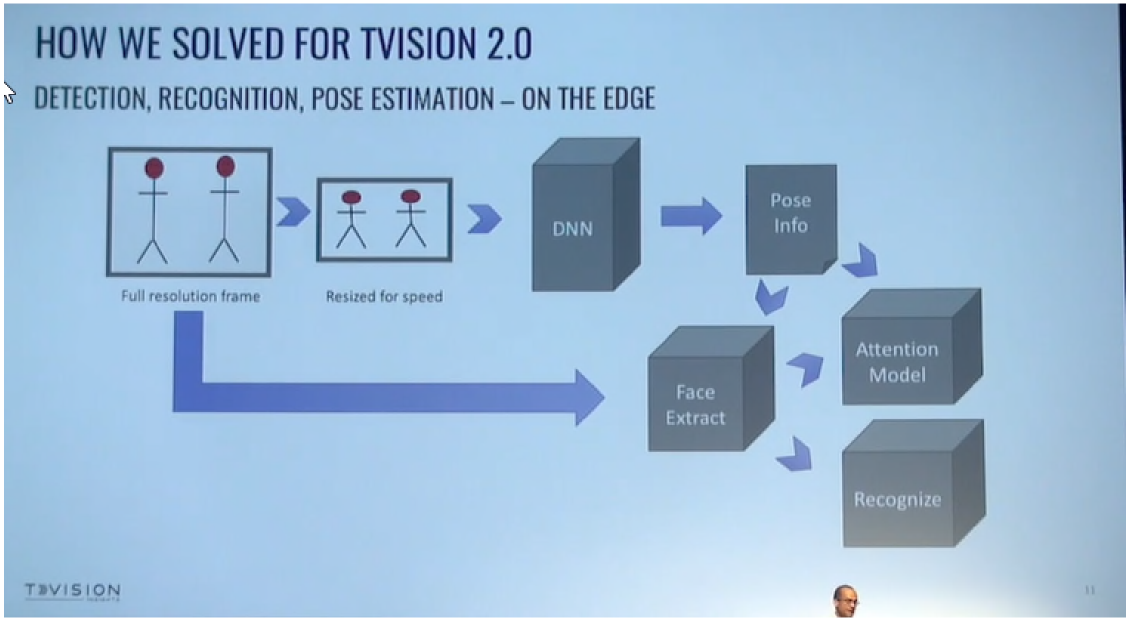
<table>
<tr><th>Element of Claim 1</th><th>TVision Insights, Inc.</th></tr>
<tr><td>determine audience identification information based on a match of the head to a known person associated with the media exposure environment; and</td><td>The TVision computer applies facial recognition to images captured by the webcam to detect who is watching the TV. (See TVision, Our Technology, https://www.tvisioninsights.com/our-technology, Complaint Ex. O; Mytvpanel, About Us, https://www.mytvpanel.com/about, Complaint Ex. P ("The system starts out in Training Mode, where it captures headshot images of your household members' faces from forehead-to-chin, ear-to-ear. . . . . . The system uses those images to tell the viewers apart and map in their demographics from your household profile."); Sidhu Interview, https://www.youtube.com/watch?v=xnFypL2JXPE at 5:32 (depicting recognition based on facial features) (screenshot below).)

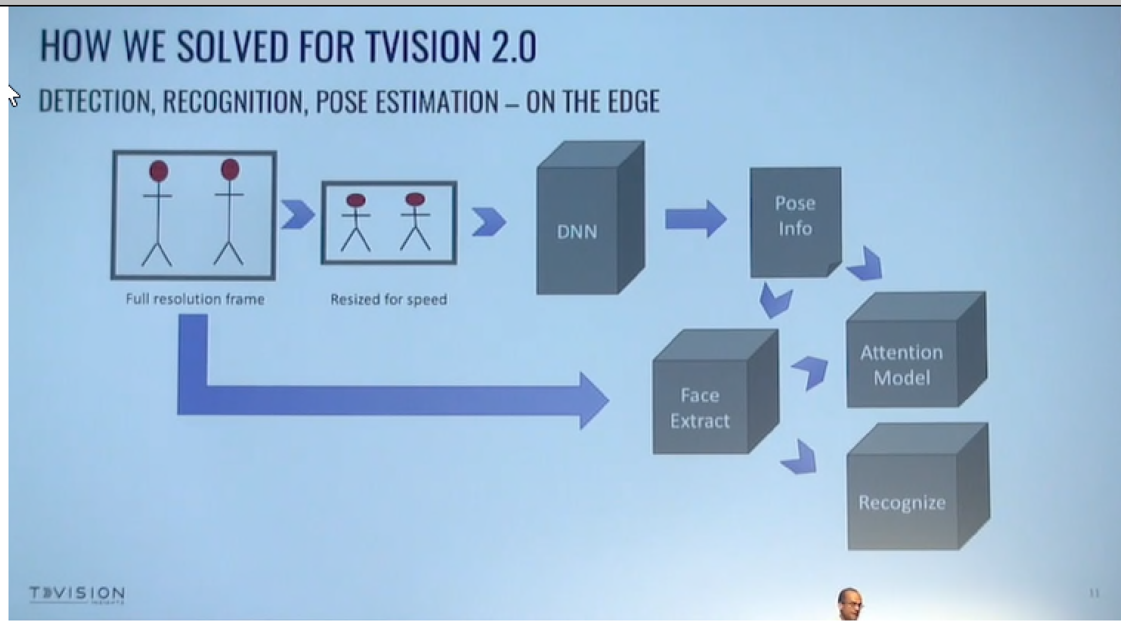

</td></tr>
</table>

Element of Claim 1	TVision Insights, Inc.
[the audience measurement system comprising:] network interface circuitry to output a signal indicative of the content identifying data and the audience identification information to a data collection facility.	The TVision computer is connected to a household's internet network using an Ethernet cable or WiFi. (*See Mytvpanel*, *About Us*, https://www.mytvpanel.com/about, Complaint Ex. P.) The TVision computer transmits attention data and the matching program or commercial name to a cloud database using network interface circuitry within the TVision computer. (TVision, "Join the TVision Panel," https://vimeo.com/295447727/3506f24b2b?embedded=true&source=video_title&owner=90679125 at 1:32 ("After completing training mode, the system can tell which person is in the room and when their eyes are on the screen. This attention data is transmitted back to TVision in a text file."); *Mytvpanel*, *About Us*, https://www.mytvpanel.com/about, Complaint Ex. P (stating that the matching program or commercial name is sent back to TVision).)

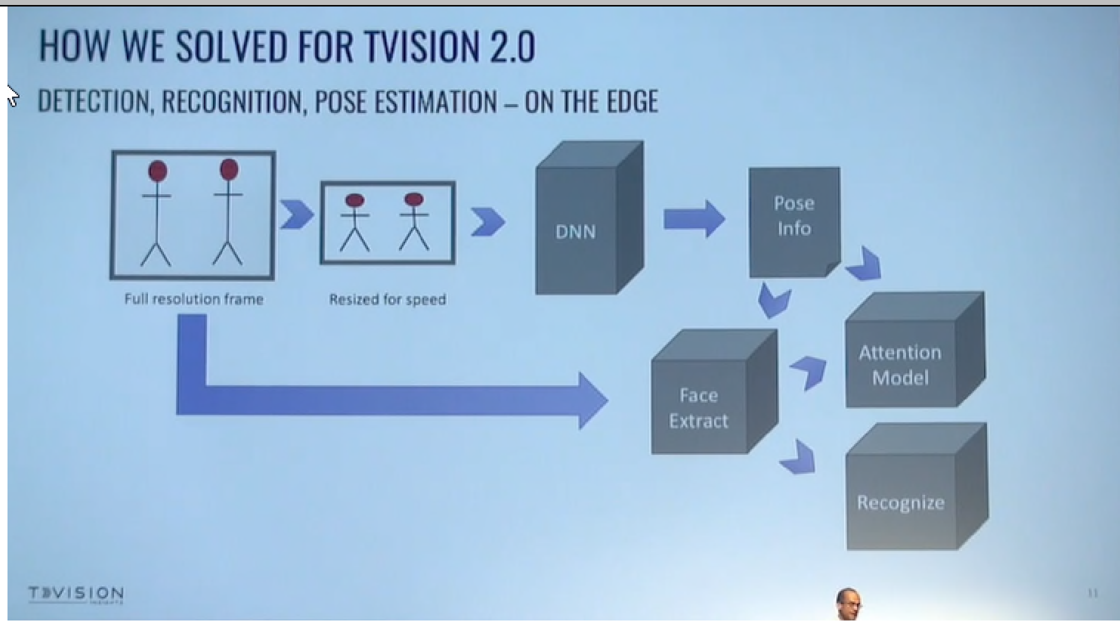
<table>
<tr><th>Element of Claim 4</th><th>TVision Insights, Inc.</th></tr>
<tr><td>4. The audience measurement system of claim 1, wherein the processor circuitry is to:

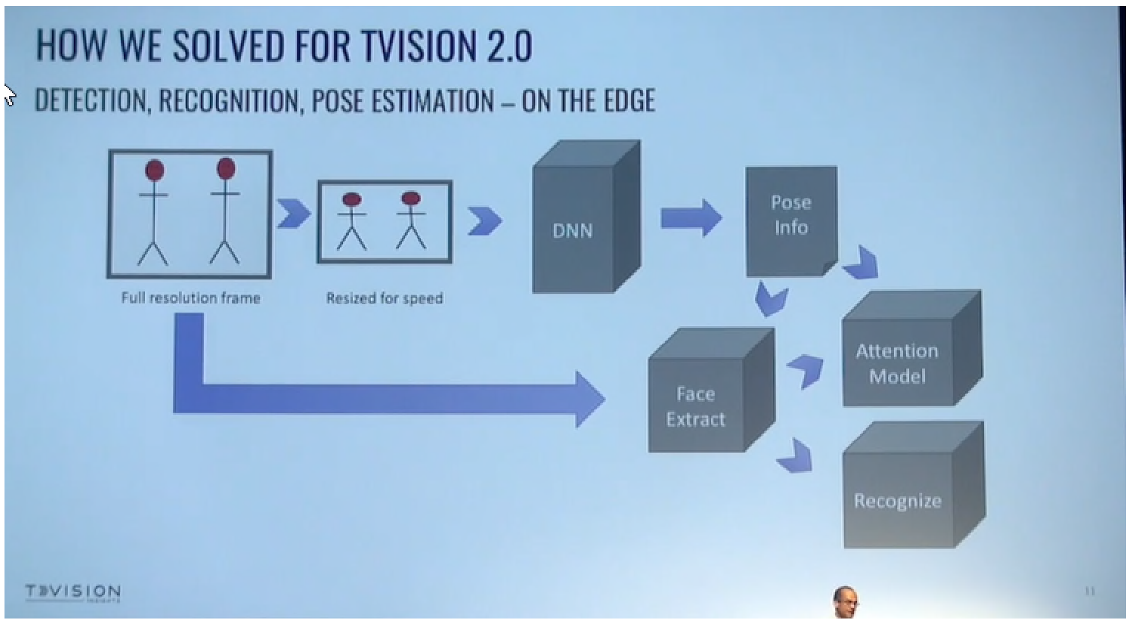
reduce a resolution of a first image of the one or more of the images of the media exposure environment to obtain a reduced-resolution image, and</td><td>The TVision computer reduces the resolution of images captured by the webcam. (See Sidhu Interview, https://www.youtube.com/watch?v=xnFypL2JXPE at 5:20-5:24 (“We have to shrink the frame before we pass it to the neural net.”) (screenshot below).)

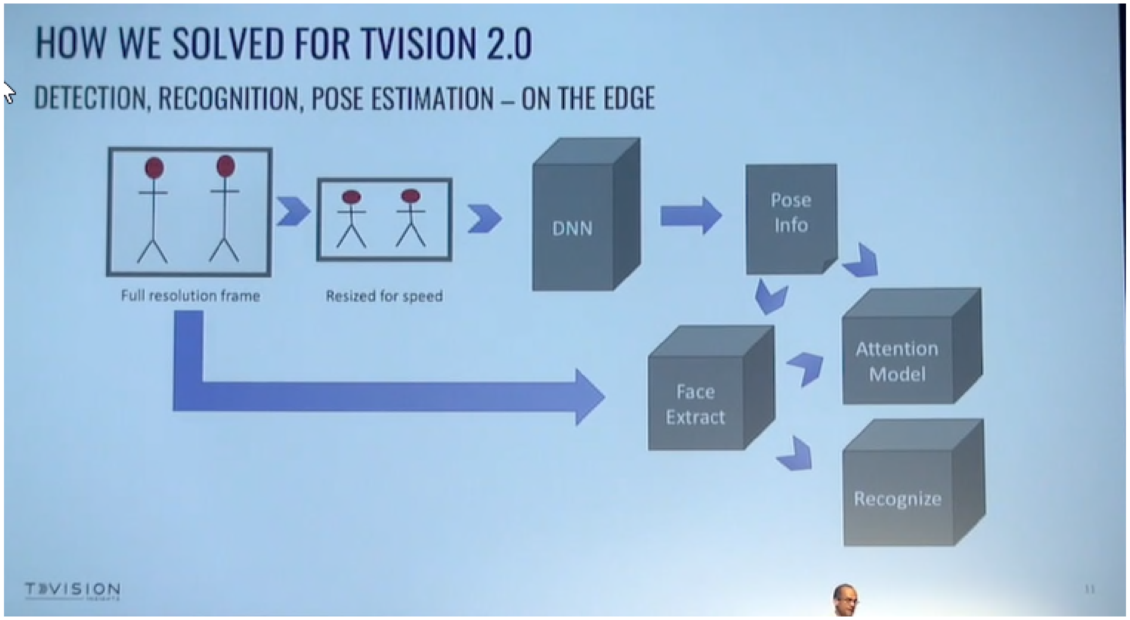

</td></tr>
<tr><td>determine the orientation of the head with respect to the camera based on the reduced-resolution image.</td><td>The TVision computer provides the resized frame to the deep neural network (“DNN”), which determines the pose information using the reduced-resolution image. The TVision computer uses the pose information as the basis to extract the head pose information. (See Sidhu Interview, https://www.youtube.com/watch?v=xnFypL2JXPE at 5:32-5:47 (“But eventually, all of the pose data makes its way to our attention measurement. And that’s where we extract out the head pose, the facial features, and all that to compute whether or not somebody is paying attention and what their gaze is.”) (screenshot below).)</td></tr>
</table>

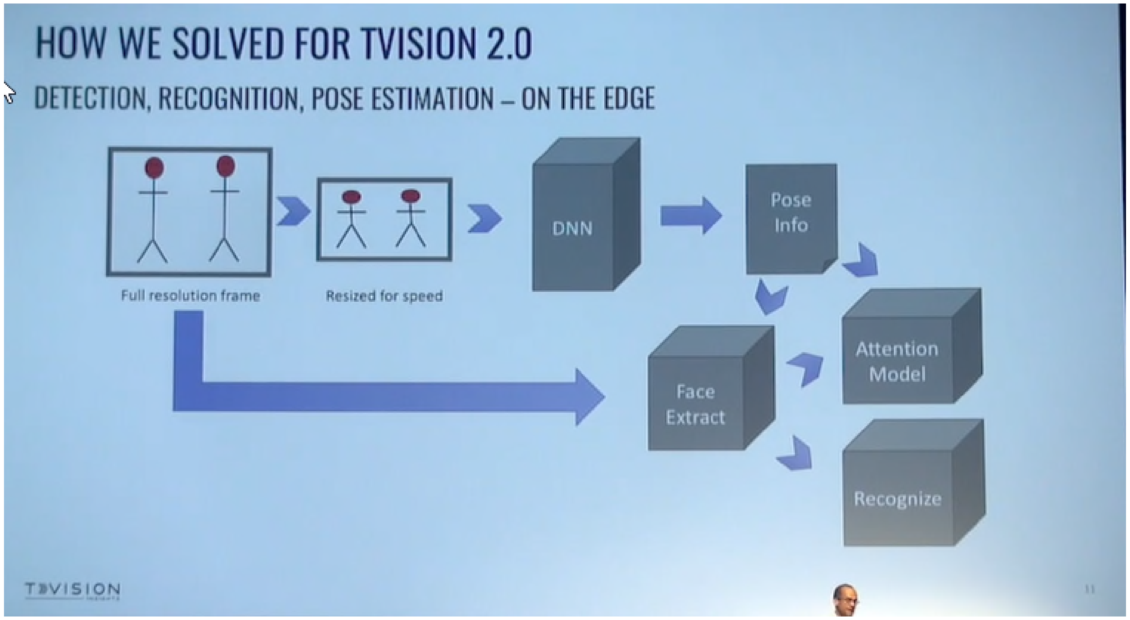
Element of Claim 4	**TVision Insights, Inc.**
	

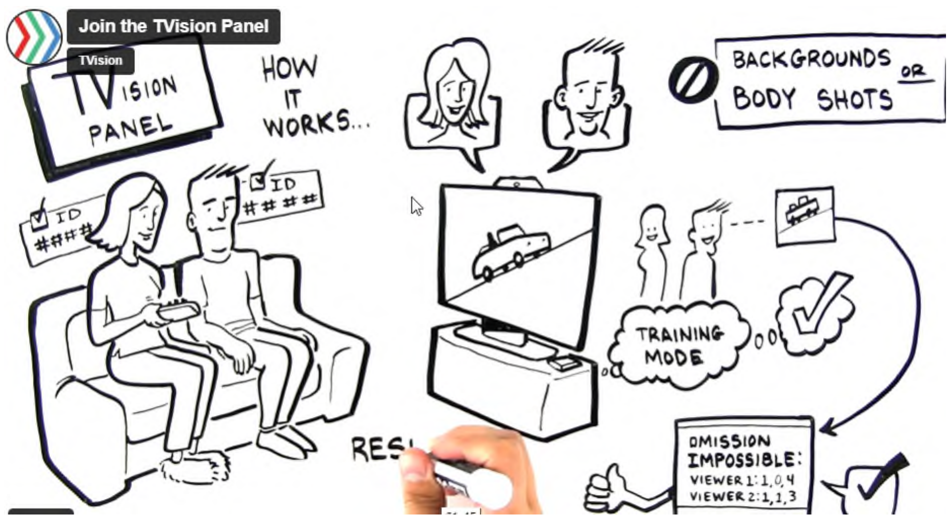
Element of Claim 5	**TVision Insights, Inc.**
5. The audience measurement system of claim 4, wherein the processor circuitry is to determine the audience identification	As explained above, the pose information output by the DNN is determined using a resized (reduced-resolution) frame. The TVision computer extracts facial signatures from a region of a full-resolution frame using the pose information output by the DNN. (*See* Sidhu Interview, https://www.youtube.com/watch?v=xnFypL2JXPE at 5:32-5:47 ("But eventually, all of the pose data makes its way to our attention measurement. And that's where we extract out the head pose, the facial features, and all that to compute whether or not somebody is paying attention and what their gaze is.") (screenshot below).)

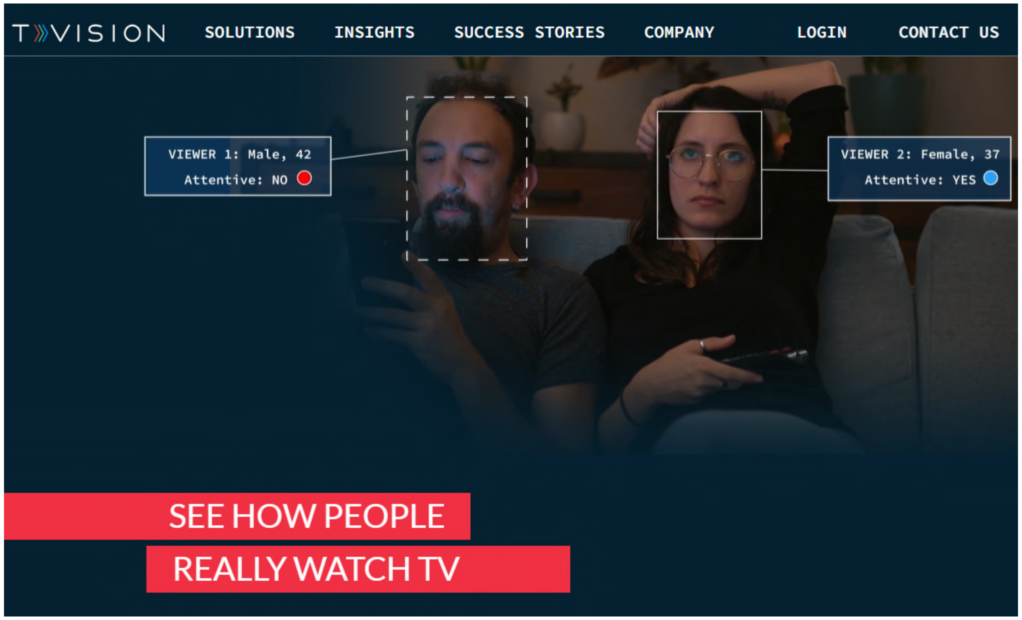
Element of Claim 5	TVision Insights, Inc.
information by: (i) generating a facial signature from a region of a second image of the one or more of the images corresponding to a location of the head in the reduced-resolution image, and (ii) comparing the generated facial signature to a database of facial signatures.	

Element of Claim 6	TVision Insights, Inc.
6. The audience measurement system of claim 4, wherein the processor circuitry is to: identify a region corresponding to the head within the first image from which the reduced-resolution image was obtained;	As explained above, the pose information output by the DNN is determined using a resized (reduced-resolution) frame. The TVision computer extracts facial signatures from a region of a full-resolution frame using the pose information output by the DNN. (*See* Sidhu Interview, https://www.youtube.com/watch?v=xnFypL2JXPE at 5:32-5:47 ("But eventually, all of the pose data makes its way to our attention measurement. And that's where we extract out the head pose, the facial features, and all that to compute whether or not somebody is paying attention and what their gaze is.") (screenshot below).) 

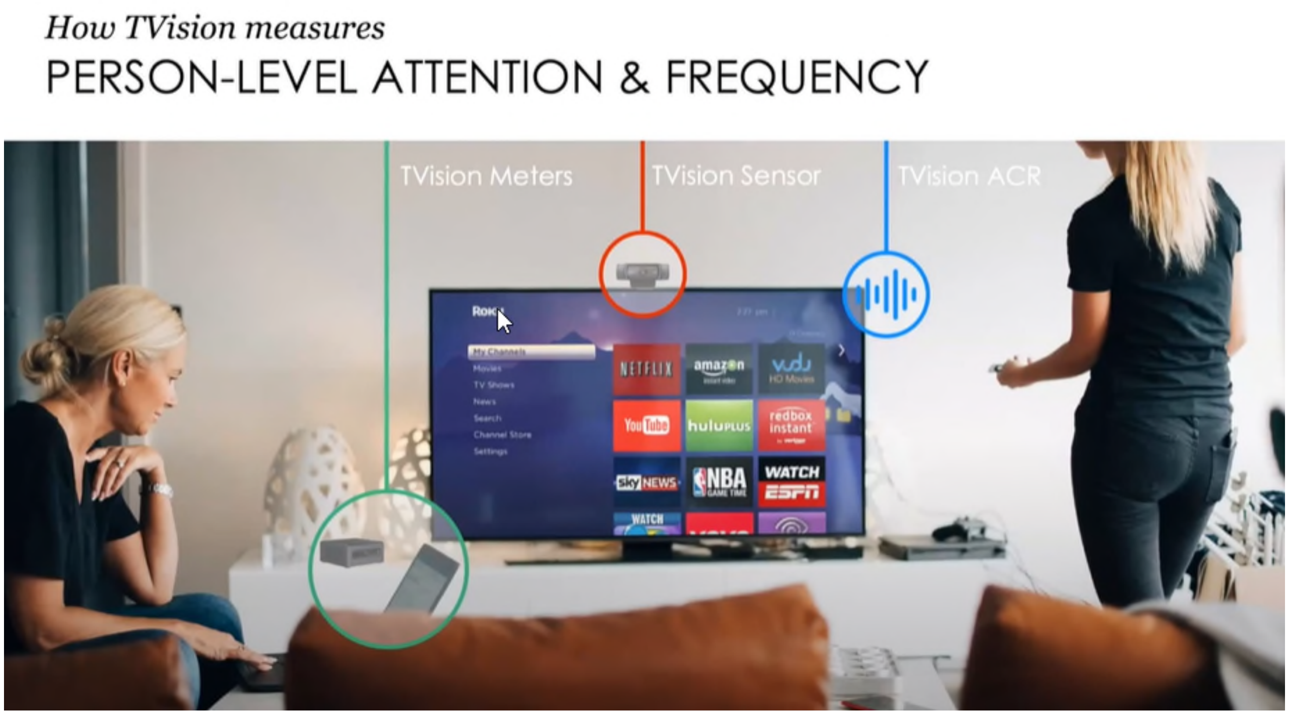
Element of Claim 6	TVision Insights, Inc.
analyze a portion corresponding to the identified region of a second image; and	As explained above, the pose information output by the DNN is determined using a resized (reduced-resolution) frame. The TVision computer extracts facial signatures from a region of a full-resolution frame using the pose information output by the DNN. (*See* Sidhu Interview, https://www.youtube.com/watch?v=xnFypL2JXPE at 5:32-5:47 ("But eventually, all of the pose data makes its way to our attention measurement. And that's where we extract out the head pose, the facial features, and all that to compute whether or not somebody is paying attention and what their gaze is.") (screenshot below).) 
based on the analysis of the region of the second image, determine that the head matches the known person.	The TVision computer applies facial recognition to images captured by the webcam to detect who is watching the TV. (*See TVision*, *Our Technology*, https://www.tvisioninsights.com/our-technology, Complaint Ex. O; *Mytvpanel*, *About Us*, https://www.mytvpanel.com/about, Complaint Ex. P ("The system starts out in Training Mode, where it captures headshot images of your household members' faces from forehead-to-chin, ear-to-ear. . . . . The system uses those images to tell the viewers apart and map in their demographics from your household profile."); Sidhu Interview, https://www.youtube.com/watch?v=xnFypL2JXPE at 5:32 (depicting recognition based on facial features) (screenshot below).)

Element of Claim 6	TVision Insights, Inc.
	

Element of Claim 8	**TVision Insights, Inc.**
8. The audience measurement system of claim 6, wherein the audience identification information includes an identifier associated with the known person and the processor circuitry is to, responsive to a determination that the head matches the known person, cause the identifier associated with the person to be stored in the memory.	The TVision computer generates records of which household members watched which content. The records include anonymized identifiers for each household member that are stored in memory. (TVision, "Join the TVision Panel," https://vimeo.com/295447727/3506f24b2b?embedded=true&source=video_title&owner=90679125 at 1:25-1:42 ("The system uses those [headshot] images to tell the viewers apart and assign an anonymized ID for each person. After completing training mode, the system can tell which person is in the room and when their eyes are on the screen. This attention data is transmitted back to TVision in a text file.") (screenshot below).) 

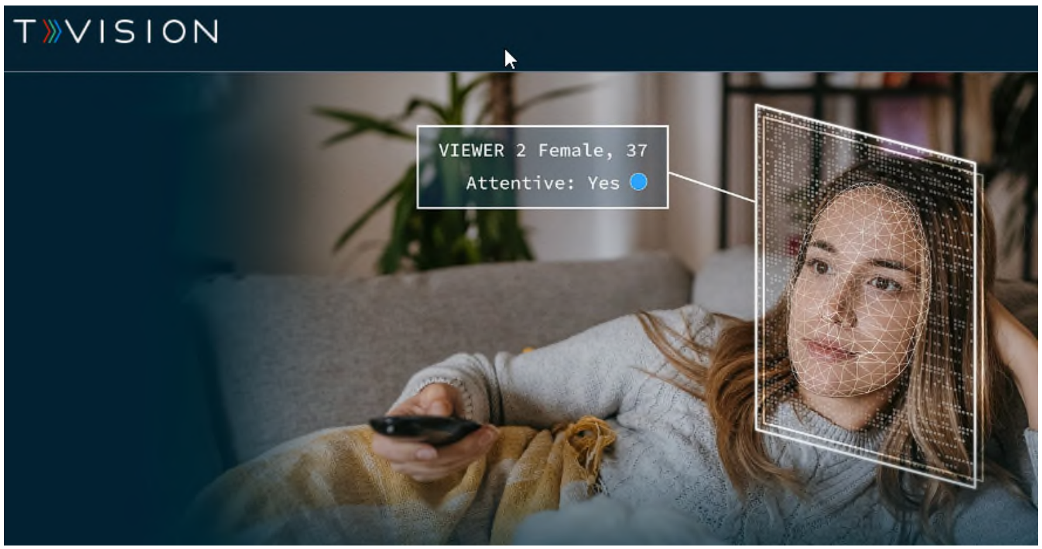
Element of Claim 9	TVision Insights, Inc.
9. An audience measurement system, comprising: an audience measurement device at a media exposure environment, the audience measurement device to:	TVision "gathers second-by-second data from a nationally representative panel of households who have signed on to help our industry understand how, what, and when they watch TV." TVision captures and reports "[w]hat program or ad is playing on the TV," "[w]hich individuals are in in the room," and "[i]f they're paying attention to the TV." (*TVision, TVision Insights*, https://www.tvisioninsights.com/, Complaint Ex. M (screenshot below).)  TVision's system includes a webcam that is set up on a user's TV (*e.g.*, placed above or below the TV with a mount clip) as well as a computer that captures audio of the program or commercial airing on the TV. (*See TVision, TVision Methodology Overview*, https://www.tvisioninsights.com/resources/tvision-methodology-overview, Complaint Ex. N; *Mytvpanel, About Us*, https://www.mytvpanel.com/about, Complaint Ex. P; *TVision, Our Technology*, https://www.tvisioninsights.com/our-technology, Complaint Ex. O; TVision, "Join the TVision Panel," https://vimeo.com/295447727/3506f24b2b?embedded=true&source=video_title&owner=90679125 at 1:09 (first screenshot below); Sidhu Interview, https://www.youtube.com/watch?v=gTBEpZo1HcM at 5:06 (second screenshot below).)

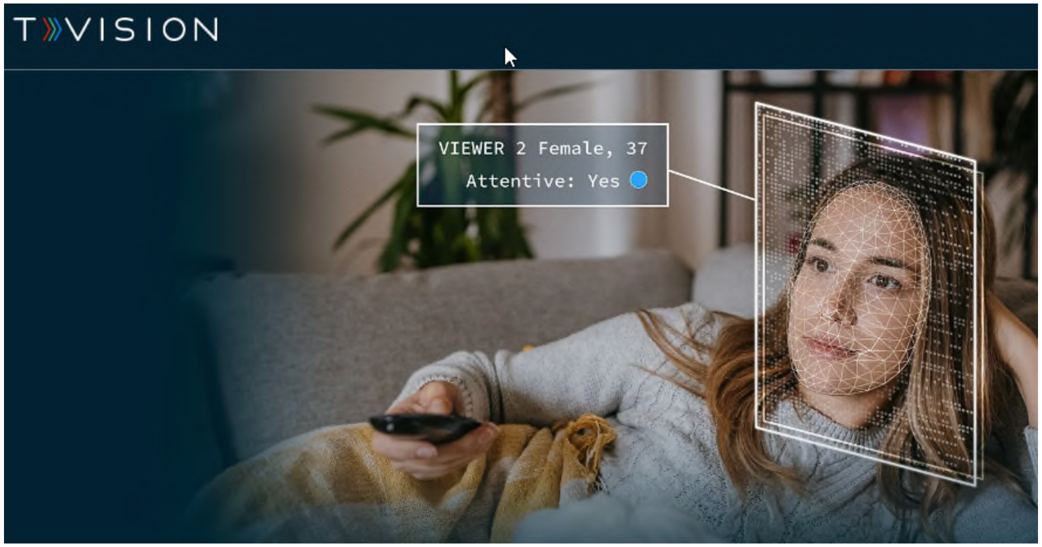
Element of Claim 9	TVision Insights, Inc.
	

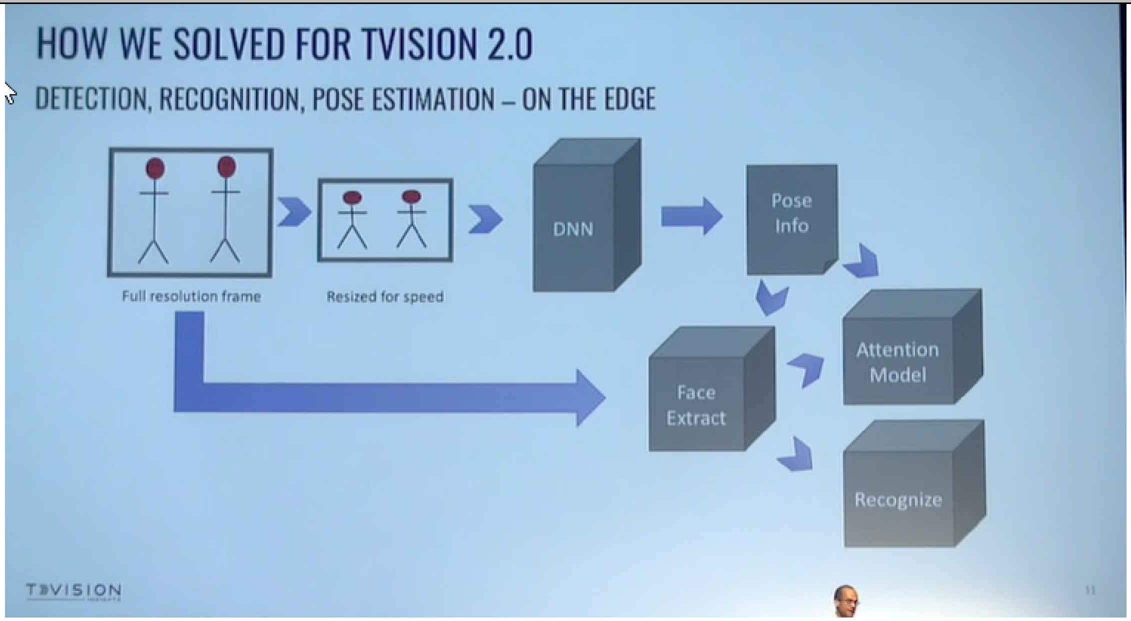
Element of Claim 9	TVision Insights, Inc.
	 The TVision computer includes a microphone array that collects audio signals output by the TV. The TVision computer includes a processor and a memory. The memory stores instructions executable by the processor to “identify the content that is playing through the TV” using the audio signals and to monitor “individual viewing behavior” using images captured by the webcam. (*See TVision, TVision Methodology Overview*, https://www.tvisioninsights.com/resources/tvision-methodology-overview, Complaint Ex. N.)

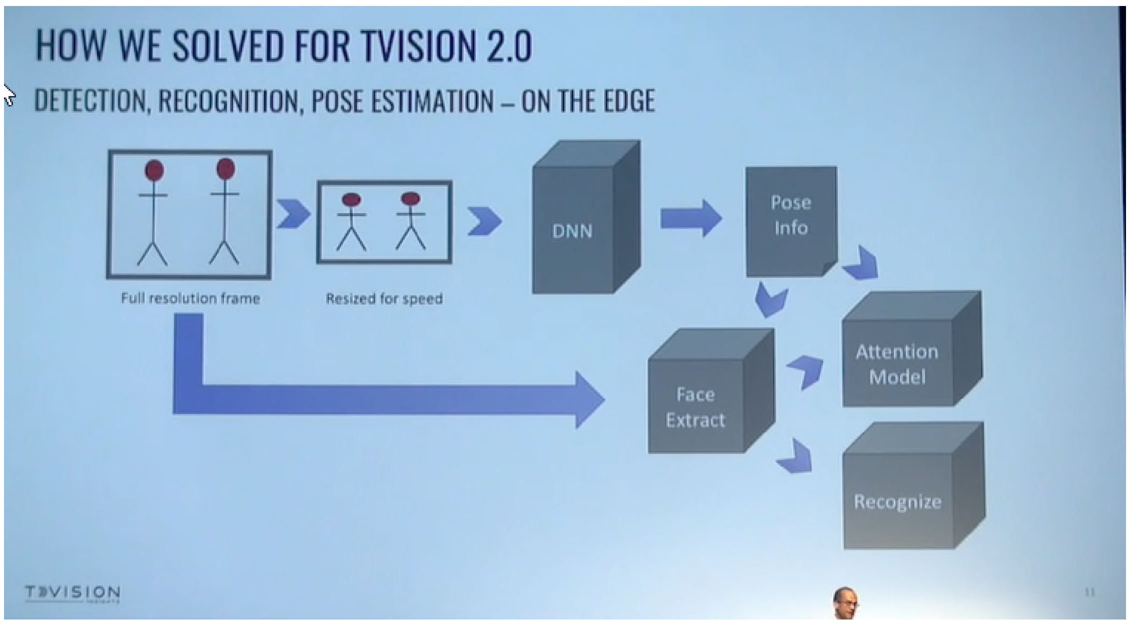
Element of Claim 9	TVision Insights, Inc.
generate an audio signature of media content presented by a television within the media exposure environment;	The TVision computer collects audio signals output by the TV and generates audio signatures (also known as fingerprints) to identify what program or commercial is on the TV. (*See TVision, TVision Methodology Overview*, https://www.tvisioninsights.com/resources/tvision-methodology-overview, Complaint Ex. N ("The microphone array on the device collects audio fingerprints that help to identify the content that is playing through the TV."); *Mytvpanel, About Us*, https://www.mytvpanel.com/about, Complaint Ex. P ("Using technology similar to 'Shazam,' our identifier detects TV audio by searching for small digital or audio tags that are unique to each program or ad. We then match those tags to shows and commercials in our database, before sending back the matching program or commercial name to TVision."); TVision, "Join the TVision Panel," https://vimeo.com/295447727/3506f24b2b?embedded=true&source=video_title&owner=90679125 at 1:48-2:00 ("Every TV program and advertisement has a unique audio fingerprint. The TVision technology scans for these unique audio fingerprints, kind of like Shazam, and is able to identify what is on the TV.").) Shazam works by generating "a digital fingerprint" of audio captured by a device. (*See Shazam, Company*, https://www.shazam.com/company, Complaint Ex. R.) To generate audio fingerprints, the TVision computer uses software provided by ACRCloud Limited ("ACRCloud"). (*See ACRCloud, Live Channel Detection*, https://www.acrcloud.com/live-channel-detection/, Complaint Ex. S; *ACRCloud, Advertising Big Data*, https://www.acrcloud.com/advertising-big-data/, Complaint Ex. T.)

Element of Claim 9	**TVision Insights, Inc.**
obtain content identifying data corresponding to the presented media content, the content identifying data based on the audio signature;	The TVision computer collects audio signals output by the TV and generates audio signatures (also known as fingerprints) to identify what program or commercial is on the TV. (*See TVision*, *TVision Methodology Overview*, https://www.tvisioninsights.com/resources/tvision-methodology-overview, Complaint Ex. N ("The microphone array on the device collects audio fingerprints that help to identify the content that is playing through the TV."); *Mytvpanel*, *About Us*, https://www.mytvpanel.com/about, Complaint Ex. P ("Using technology similar to 'Shazam,' our identifier detects TV audio by searching for small digital or audio tags that are unique to each program or ad. We then match those tags to shows and commercials in our database, before sending back the matching program or commercial name to TVision."); TVision, "Join the TVision Panel," https://vimeo.com/295447727/3506f24b2b?embedded=true&source=video_title&owner=90679125 at 1:48-2:00 ("Every TV program and advertisement has a unique audio fingerprint. The TVision technology scans for these unique audio fingerprints, kind of like Shazam, and is able to identify what is on the TV.").) Shazam works by generating "a digital fingerprint" of audio captured by a device. (*See Shazam*, *Company*, https://www.shazam.com/company, Complaint Ex. R.) To generate audio fingerprints, the TVision computer uses software provided by ACRCloud Limited ("ACRCloud"). (*See ACRCloud*, *Live Channel Detection*, https://www.acrcloud.com/live-channel-detection/, Complaint Ex. S; *ACRCloud*, *Advertising Big Data*, https://www.acrcloud.com/advertising-big-data/, Complaint Ex. T.) The TVision computer uses the above-described audio signatures to identify "the matching program or commercial name." (*See Mytvpanel*, *About Us*, https://www.mytvpanel.com/about, Complaint Ex. P.)

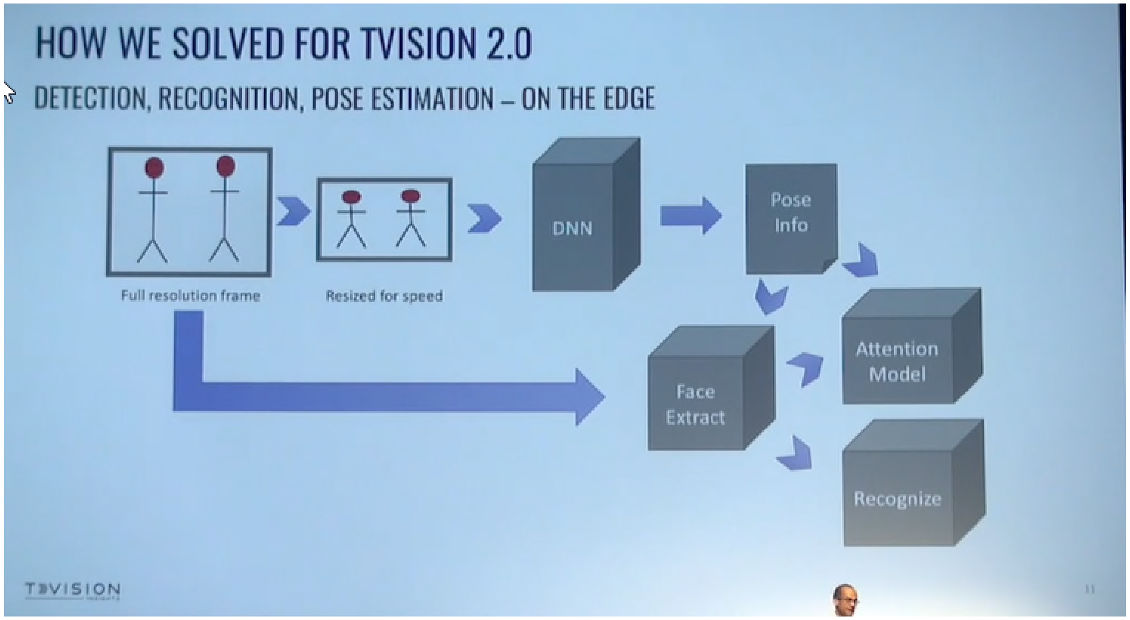
Element of Claim 9	**TVision Insights, Inc.**
while the media content corresponding to the content identifying data is presented by the television, collect a first image of the media exposure environment with a camera;	As explained above, the TVision computer monitors individual viewing behavior using images captured by the webcam mounted to the TV. The TVision computer then uses the images captured by the webcam to detect a user's head appearing in one or more of the images. (*See TVision*, *Our Technology*, https://www.tvisioninsights.com/our-technology, Complaint Ex. O (screenshot below); Sidhu Interview https://www.youtube.com/watch?v=xnFypL2JXPE at 3:10-3:20, 4:47-5:06, and 5:32-5:47 ("It is key to us to understand the pose information . . . of the audience. . . . We use deep neural nets to get pose information. We detect humans. We track humans. We kind of have trained our own model for tracking. We use something called graph cut algorithms to track the path of people across multiple frames.").) 

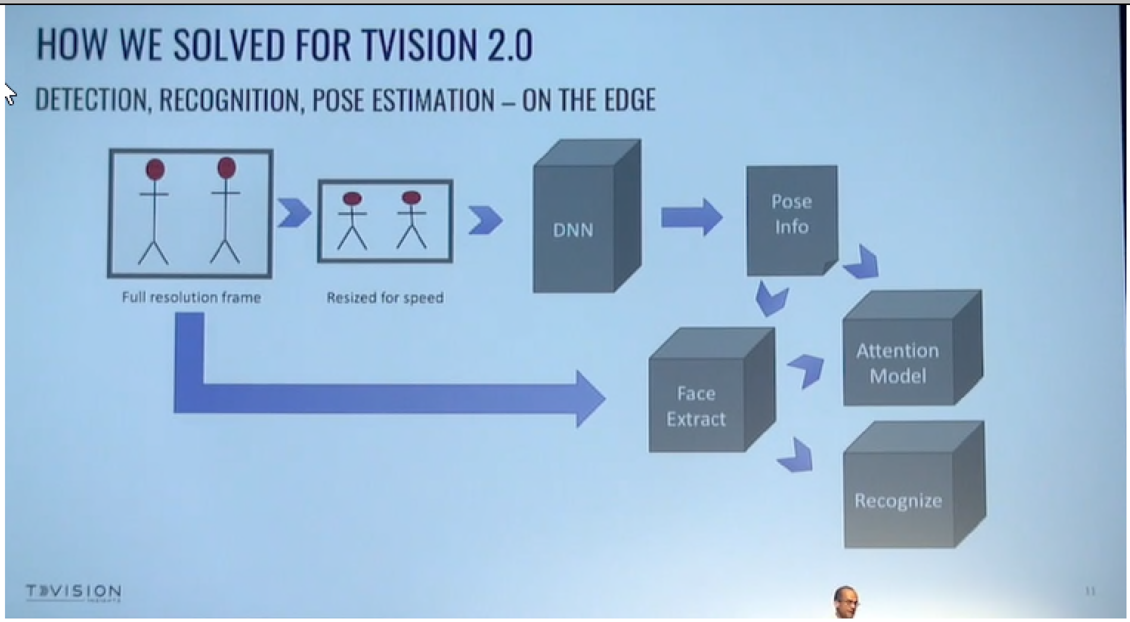
Element of Claim 9	TVision Insights, Inc.
attempt to detect a head in the first image;	The TVision computer uses the image captured by the webcam to detect a user's head appearing in the image. (*See TVision, Our Technology*, https://www.tvisioninsights.com/our-technology, Complaint Ex. O (screenshot below); Sidhu Interview, https://www.youtube.com/watch?v=xnFypL2JXPE at 3:10-3:20, 4:47-5:06, and 5:32-5:47 ("[i]t is key to us to understand the pose information . . . of the audience. . . . We use deep neural nets to get pose information. We detect humans. We track humans. We kind of have trained our own model for tracking. We use something called graph cut algorithms to track the path of people across multiple frames.).)  Source: https://www.tvisioninsights.com/our-technology (last accessed July 2022).
determine an orientation of the head with respect to the camera;	The TVision computer uses computer-vision technology to determine a head pose of the detected head. (*See* Sidhu Interview, https://www.youtube.com/watch?v=xnFypL2JXPE at 5:32-5:47 ("But eventually, all of the pose data makes its way to our attention measurement. And that's where we extract out the head pose, the facial features, and all that to compute whether or not somebody is paying attention and what their gaze is.") (screenshot below).)

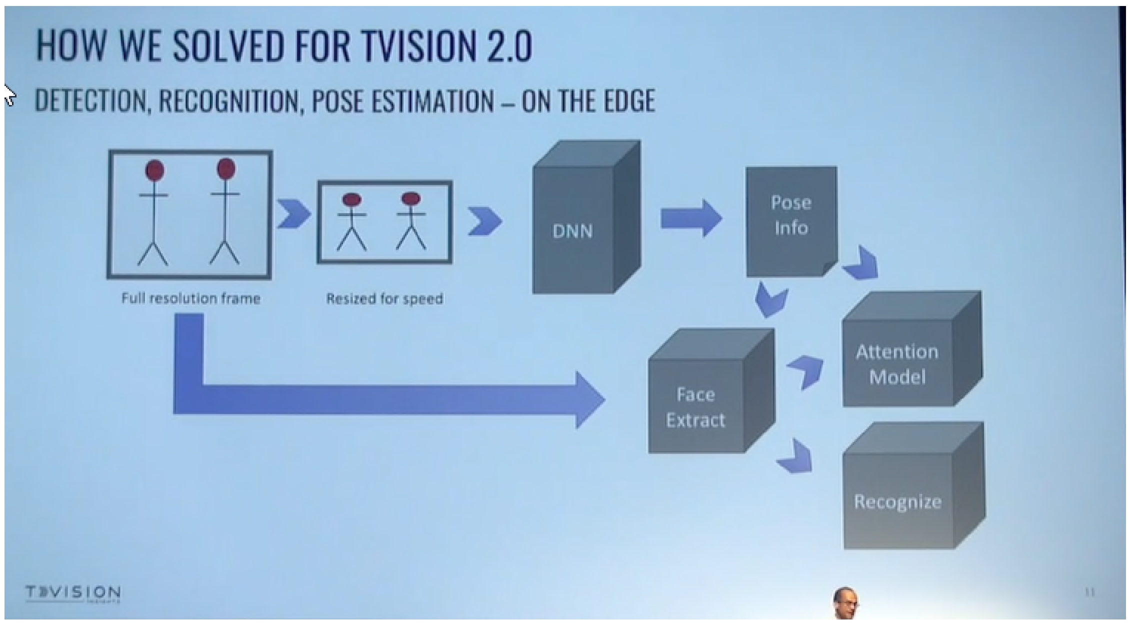
Element of Claim 9	TVision Insights, Inc.
	

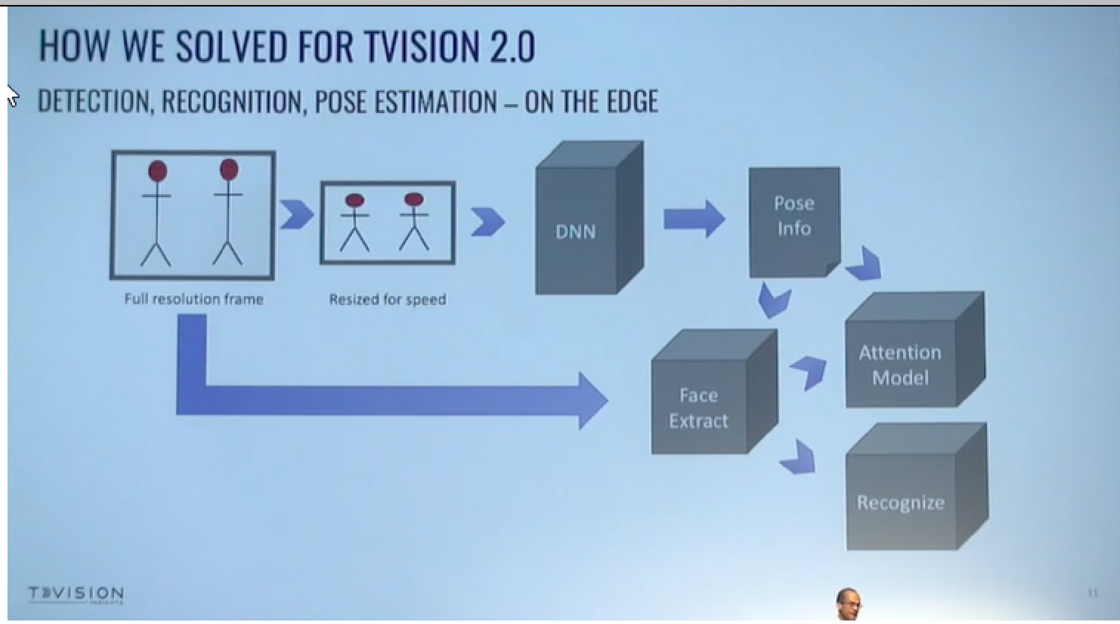
Element of Claim 9	**TVision Insights, Inc.**
determine audience identification information based on an indication that the head matches a known person associated with the audience measurement system; and	The TVision computer applies facial recognition to images captured by the webcam to detect who is watching the TV. (*See TVision, Our Technology*, https://www.tvisioninsights.com/our-technology, Complaint Ex. O; *Mytvpanel*, *About Us*, https://www.mytvpanel.com/about, Complaint Ex. P ("The system starts out in Training Mode, where it captures headshot images of your household members' faces from forehead-to-chin, ear-to-ear. . . . . . The system uses those images to tell the viewers apart and map in their demographics from your household profile."); Sidhu Interview, https://www.youtube.com/watch?v=xnFypL2JXPE at 5:32 (depicting recognition based on facial features) (screenshot below).) 

Element of Claim 9	TVision Insights, Inc.
[the audience measurement system comprising:] a data collection facility to obtain the content identifying data and the audience identification information from the audience measurement device.	The TVision computer is connected to a household's internet network using an Ethernet cable or WiFi. (*See Mytvpanel*, *About Us*, https://www.mytvpanel.com/about, Complaint Ex. P.) The TVision computer transmits attention data and the matching program or commercial name to a cloud database using network interface circuitry within the TVision computer. (TVision, "Join the TVision Panel," https://vimeo.com/295447727/3506f24b2b?embedded=true&source=video_title&owner=90679125 at 1:32 ("After completing training mode, the system can tell which person is in the room and when their eyes are on the screen. This attention data is transmitted back to TVision in a text file."); *Mytvpanel*, *About Us*, https://www.mytvpanel.com/about, Complaint Ex. P (stating that the matching program or commercial name is sent back to TVision).)

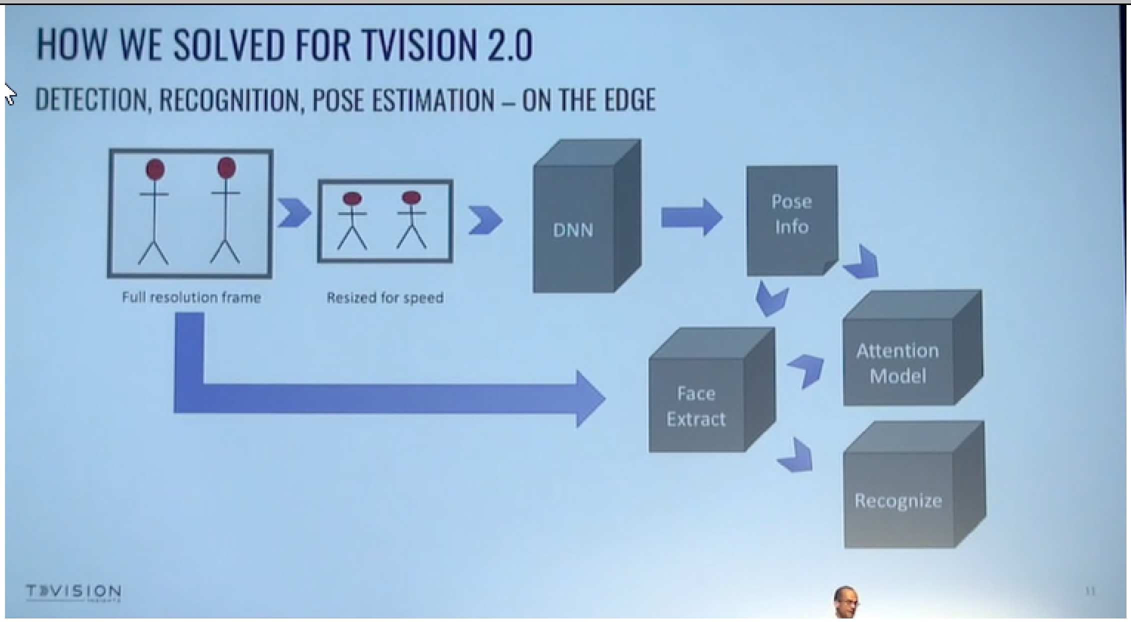
Element of Claim 11	TVision Insights, Inc.
11. The audience measurement system of claim 9, wherein the audience measurement device is to: reduce a resolution of the first image of the media exposure environment to obtain a reduced-resolution image, and	The TVision computer reduces the resolution of images captured by the webcam. (*See* Sidhu Interview, https://www.youtube.com/watch?v=xnFypL2JXPE at 5:20-5:24 ("We have to shrink the frame before we pass it to the neural net.") (screenshot below).) 
determine the orientation of the head with respect to the camera based on the reduced-resolution image.	The TVision computer provides the resized frame to the deep neural network ("DNN"), which determines the pose information using the reduced-resolution image. The TVision computer uses the pose information as the basis to extract the head pose information. (*See* Sidhu Interview, https://www.youtube.com/watch?v=xnFypL2JXPE at 5:32-5:47 ("But eventually, all of the pose data makes its way to our attention measurement. And that's where we extract out the head pose, the facial features, and all that to compute whether or not somebody is paying attention and what their gaze is.") (screenshot below).)

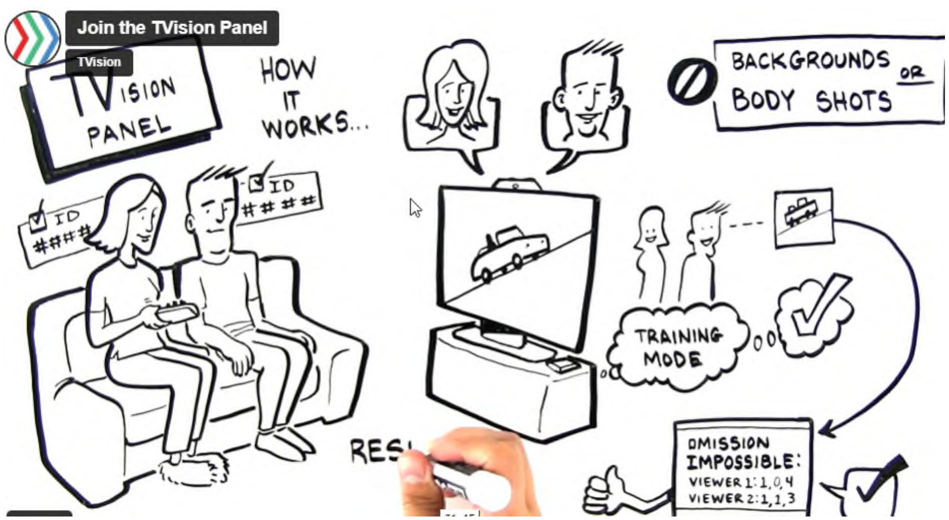
Element of Claim 11	TVision Insights, Inc.
	

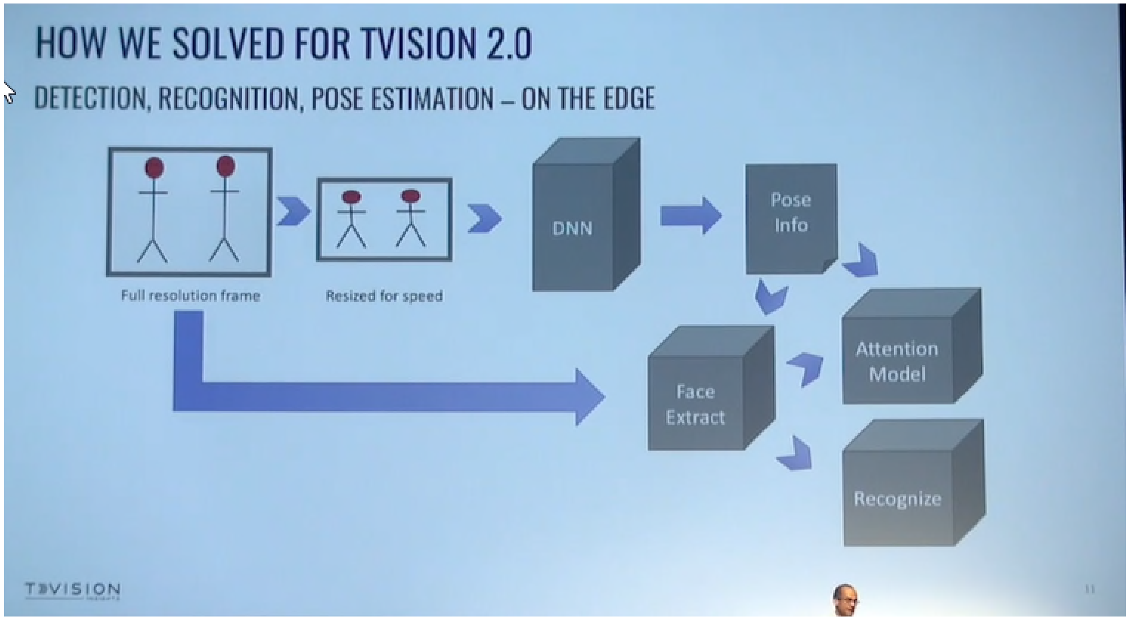
Element of Claim 12	**TVision Insights, Inc.**
12. The audience measurement system of claim 11, wherein the audience measurement device is to: identify a region corresponding to the head within the first image from which the reduced-resolution image was obtained; and	As explained above, the pose information output by the DNN is determined using a resized (reduced-resolution) frame. The TVision computer extracts facial signatures from a region of a full-resolution frame using the pose information output by the DNN. (*See* Sidhu Interview, https://www.youtube.com/watch?v=xnFypL2JXPE at 5:32-5:47 ("But eventually, all of the pose data makes its way to our attention measurement. And that's where we extract out the head pose, the facial features, and all that to compute whether or not somebody is paying attention and what their gaze is.") (screenshot below).) 
analyze a portion of a second image corresponding to the identified region to determine whether the head matches the known person.	As explained above, the pose information output by the DNN is determined using a resized (reduced-resolution) frame. The TVision computer extracts facial signatures from a region of a full-resolution frame using the pose information output by the DNN. (*See* Sidhu Interview, https://www.youtube.com/watch?v=xnFypL2JXPE at 5:32-5:47 ("But eventually, all of the pose data makes its way to our attention measurement. And that's where we extract out the head pose, the facial features, and all that to compute whether or not somebody is paying attention and what their gaze is.") (screenshot below).)

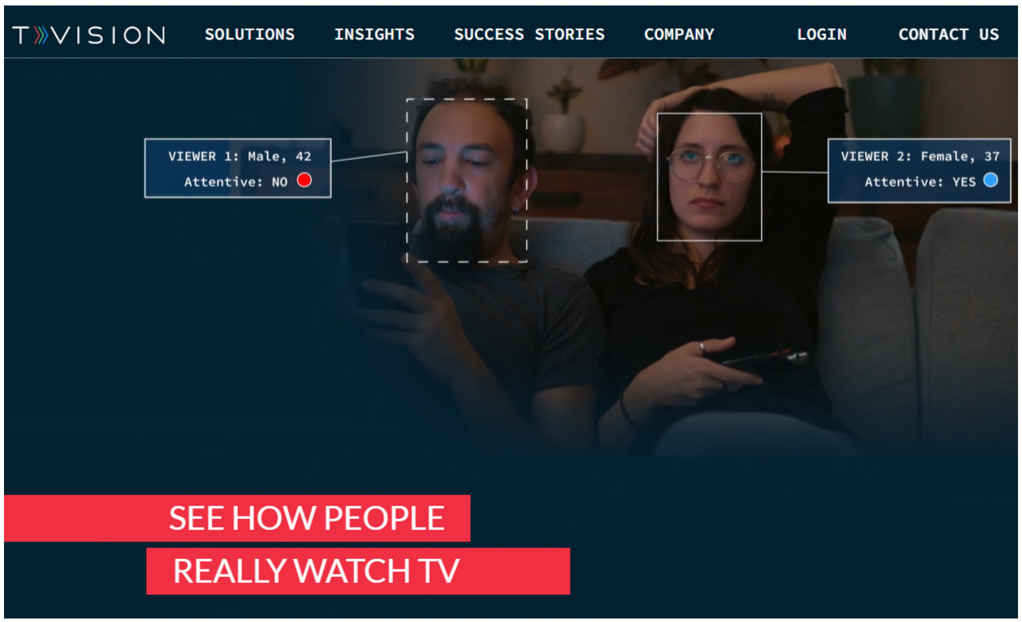
<table>
<tr><th>Element of Claim 12</th><th>TVision Insights, Inc.</th></tr>
<tr><td></td><td>

The TVision computer applies facial recognition to images captured by the webcam to detect who is watching the TV. (See TVision, Our Technology, https://www.tvisioninsights.com/our-technology, Complaint Ex. O; Mytvpanel, About Us, https://www.mytvpanel.com/about, Complaint Ex. P ("The system starts out in Training Mode, where it captures headshot images of your household members' faces from forehead-to-chin, ear-to-ear. . . . . . The system uses those images to tell the viewers apart and map in their demographics from your household profile."); Sidhu Interview, https://www.youtube.com/watch?v=xnFypL2JXPE at 5:32 (depicting recognition based on facial features) (screenshot below).)</td></tr>
</table>

Element of Claim 12	TVision Insights, Inc.
	

Element of Claim 13	TVision Insights, Inc.
13. The audience measurement system of claim 12, wherein the audience identification information includes an identifier associated with the known person and the audience measurement device is to, responsive to the indication that the head matches the known person, cause recordation of the identifier associated with the known person in a memory.	The TVision computer generates records of which household members watched which content. The records include anonymized identifiers for each household member that are stored in memory. (TVision, "Join the TVision Panel," https://vimeo.com/295447727/3506f24b2b?embedded=true&source=video_title&owner=90679125 at 1:25-1:42 ("The system uses those [headshot] images to tell the viewers apart and assign an anonymized ID for each person. After completing training mode, the system can tell which person is in the room and when their eyes are on the screen. This attention data is transmitted back to TVision in a text file.") (screenshot below).) 

<table>
<tr><th>Element of Claim 14</th><th>TVision Insights, Inc.</th></tr>
<tr><td>14. The audience measurement system of claim 11, wherein the audience measurement device is to determine the audience identification information by: (i) generating a facial signature from a region of a second image corresponding to a location of the head in the reduced-resolution image, and (ii) comparing the generated facial signature to a database of facial signatures.</td><td>As explained above, the pose information output by the DNN is determined using a resized (reduced-resolution) frame. The TVision computer extracts facial signatures from a region of a full-resolution frame using the pose information output by the DNN. (See Sidhu Interview, https://www.youtube.com/watch?v=xnFypL2JXPE at 5:32-5:47 (“But eventually, all of the pose data makes its way to our attention measurement. And that’s where we extract out the head pose, the facial features, and all that to compute whether or not somebody is paying attention and what their gaze is.”) (screenshot below).)

</td></tr>
</table>

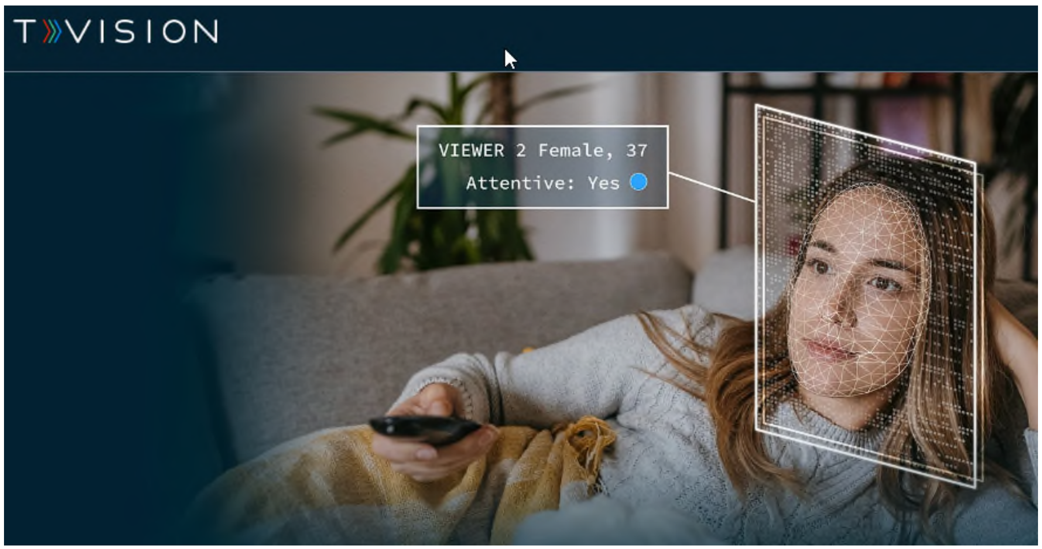
Element of Claim 16	TVision Insights, Inc.
16. A method for obtaining exposure data for a media exposure environment, the method comprising:	TVision “gathers second-by-second data from a nationally representative panel of households who have signed on to help our industry understand how, what, and when they watch TV.” TVision captures and reports “[w]hat program or ad is playing on the TV,” “[w]hich individuals are in in the room,” and “[i]f they’re paying attention to the TV.” (*TVision, TVision Insights*, https://www.tvisioninsights.com/, Complaint Ex. M (screenshot below).)  TVision’s system includes a webcam that is set up on a user’s TV (*e.g.*, placed above or below the TV with a mount clip) as well as a computer that captures audio of the program or commercial airing on the TV. (*See TVision, TVision Methodology Overview*, https://www.tvisioninsights.com/resources/tvision-methodology-overview, Complaint Ex. N; *Mytvpanel, About Us*, https://www.mytvpanel.com/about, Complaint Ex. P; *TVision, Our Technology*, https://www.tvisioninsights.com/our-technology, Complaint Ex. O; TVision, “Join the TVision Panel,” https://vimeo.com/295447727/3506f24b2b?embedded=true&source=video_title&owner=90679125 at 1:09 (first screenshot below); Sidhu Interview, https://www.youtube.com/watch?v=gTBEpZo1HcM at 5:06 (second screenshot below).)

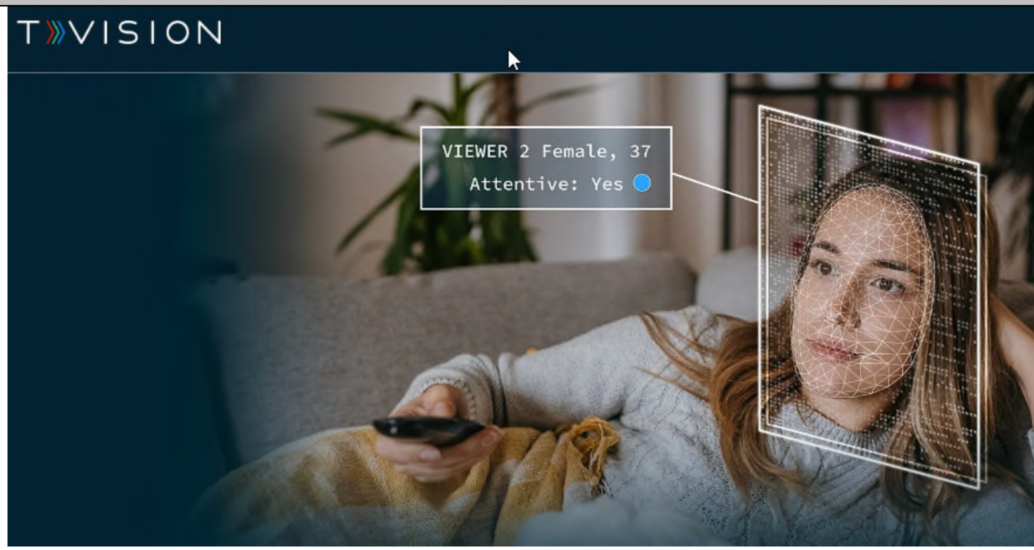
Element of Claim 16	TVision Insights, Inc.
	

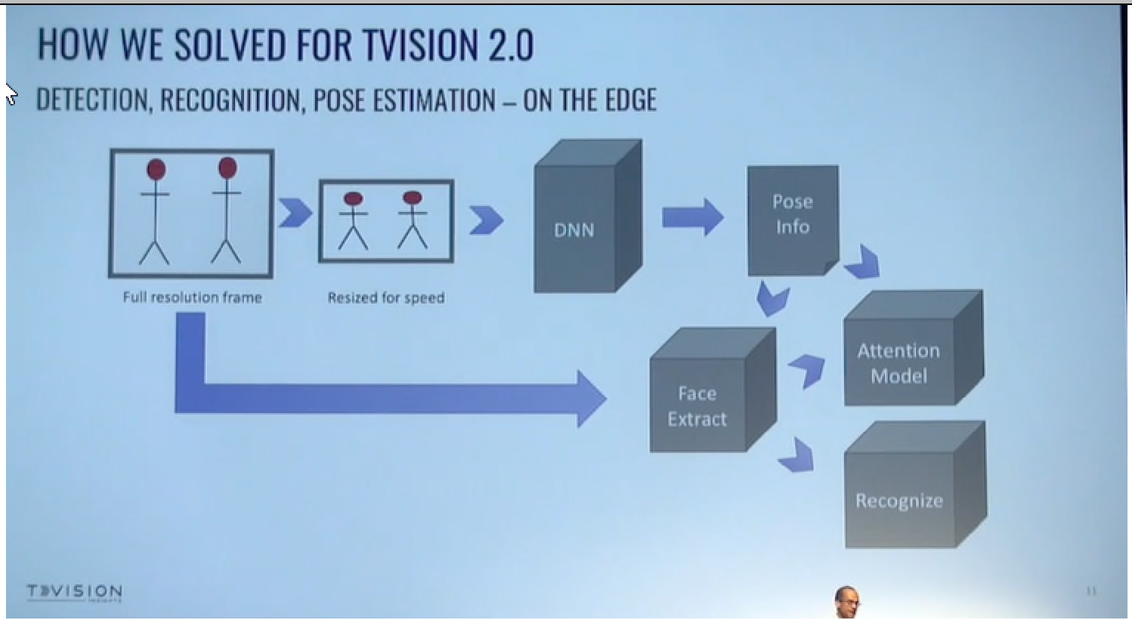
Element of Claim 16	TVision Insights, Inc.
	How TVision measures PERSON-LEVEL ATTENTION & FREQUENCY  The TVision computer includes a microphone array that collects audio signals output by the TV. The TVision computer includes a processor and a memory. The memory stores instructions executable by the processor to "identify the content that is playing through the TV" using the audio signals and to monitor "individual viewing behavior" using images captured by the webcam. (*See TVision*, *TVision Methodology Overview*, https://www.tvisioninsights.com/resources/tvision-methodology-overview, Complaint Ex. N.)

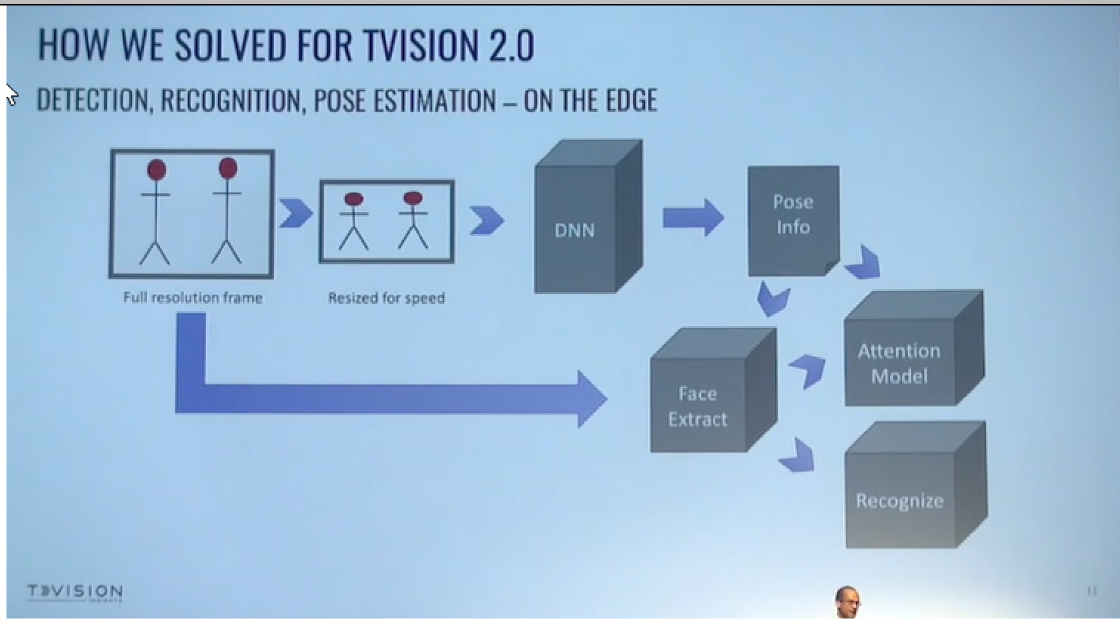
Element of Claim 16	TVision Insights, Inc.
generating, using a program detector, an audio signature of media content presented by a television within the media exposure environment;	The TVision computer collects audio signals output by the TV and generates audio signatures (also known as fingerprints) to identify what program or commercial is on the TV. (*See TVision, TVision Methodology Overview*, https://www.tvisioninsights.com/resources/tvision-methodology-overview, Complaint Ex. N ("The microphone array on the device collects audio fingerprints that help to identify the content that is playing through the TV."); *Mytvpanel, About Us*, https://www.mytvpanel.com/about, Complaint Ex. P ("Using technology similar to 'Shazam,' our identifier detects TV audio by searching for small digital or audio tags that are unique to each program or ad. We then match those tags to shows and commercials in our database, before sending back the matching program or commercial name to TVision."); TVision, "Join the TVision Panel," https://vimeo.com/295447727/3506f24b2b?embedded=true&source=video_title&owner=90679125 at 1:48-2:00 ("Every TV program and advertisement has a unique audio fingerprint. The TVision technology scans for these unique audio fingerprints, kind of like Shazam, and is able to identify what is on the TV.").) Shazam works by generating "a digital fingerprint" of audio captured by a device. (*See Shazam, Company*, https://www.shazam.com/company, Complaint Ex. R.) To generate audio fingerprints, the TVision computer uses software provided by ACRCloud Limited ("ACRCloud"). (*See ACRCloud, Live Channel Detection*, https://www.acrcloud.com/live-channel-detection/, Complaint Ex. S; *ACRCloud, Advertising Big Data*, https://www.acrcloud.com/advertising-big-data/, Complaint Ex. T.)

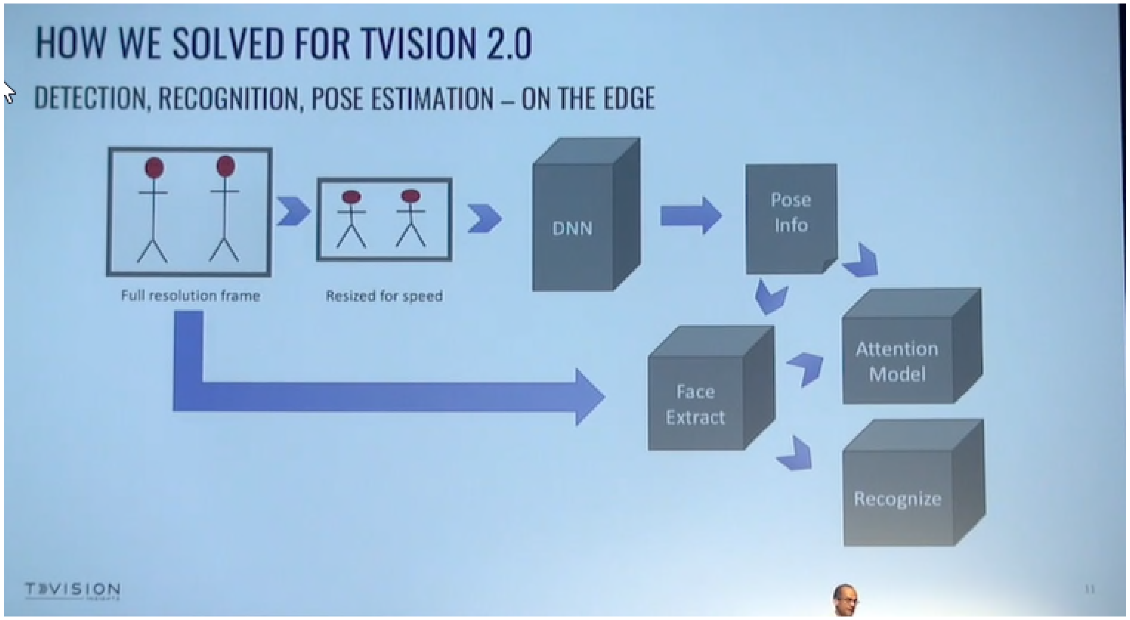
Element of Claim 16	**TVision Insights, Inc.**
obtaining, based on the audio signature, content identifying data corresponding to the presented media content;	The TVision computer collects audio signals output by the TV and generates audio signatures (also known as fingerprints) to identify what program or commercial is on the TV. (*See TVision*, *TVision Methodology Overview*, https://www.tvisioninsights.com/resources/tvision-methodology-overview, Complaint Ex. N ("The microphone array on the device collects audio fingerprints that help to identify the content that is playing through the TV."); *Mytvpanel, About Us*, https://www.mytvpanel.com/about, Complaint Ex. P ("Using technology similar to 'Shazam,' our identifier detects TV audio by searching for small digital or audio tags that are unique to each program or ad. We then match those tags to shows and commercials in our database, before sending back the matching program or commercial name to TVision."); TVision, "Join the TVision Panel," https://vimeo.com/295447727/3506f24b2b?embedded=true&source=video_title&owner=90679125 at 1:48-2:00 ("Every TV program and advertisement has a unique audio fingerprint. The TVision technology scans for these unique audio fingerprints, kind of like Shazam, and is able to identify what is on the TV.").) Shazam works by generating "a digital fingerprint" of audio captured by a device. (*See Shazam*, *Company*, https://www.shazam.com/company, Complaint Ex. R.) To generate audio fingerprints, the TVision computer uses software provided by ACRCloud Limited ("ACRCloud"). (*See ACRCloud*, *Live Channel Detection*, https://www.acrcloud.com/live-channel-detection/, Complaint Ex. S; *ACRCloud*, *Advertising Big Data*, https://www.acrcloud.com/advertising-big-data/, Complaint Ex. T.) The TVision computer uses the above-described audio signatures to identify "the matching program or commercial name." (*See Mytvpanel*, *About Us*, https://www.mytvpanel.com/about, Complaint Ex. P.)

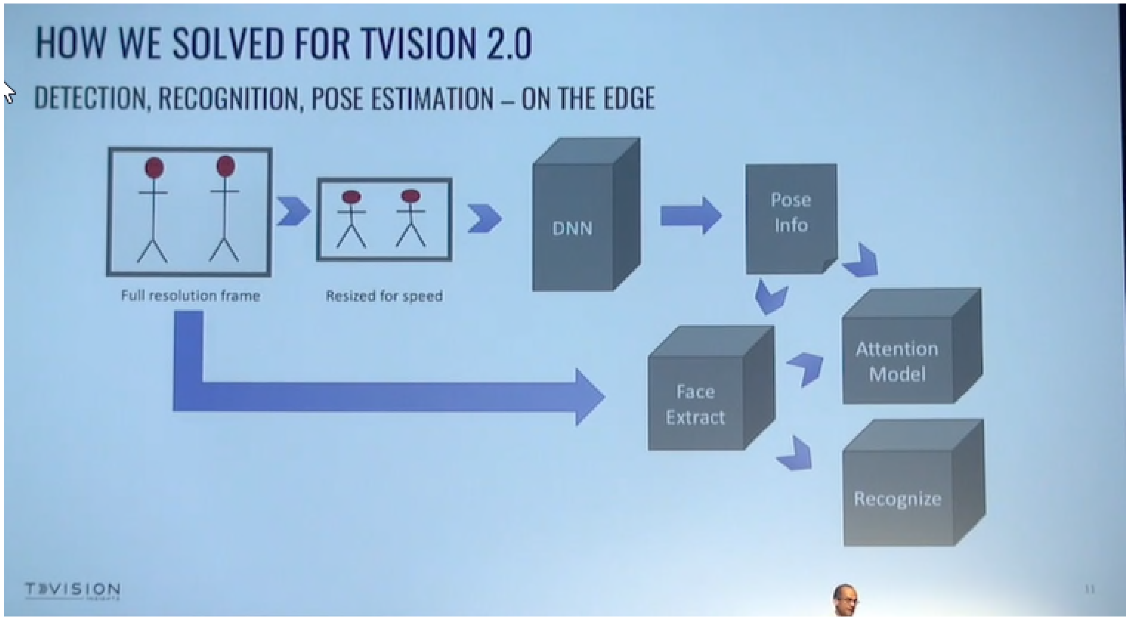
Element of Claim 16	TVision Insights, Inc.
while the media content corresponding to the content identifying data is presented by the television, capturing first and second images of the media exposure environment with a camera;	As explained above, the TVision computer monitors individual viewing behavior using images captured by the webcam mounted to the TV. The TVision computer then uses the images captured by the webcam to detect a user's head appearing in one or more of the images. (*See TVision, Our Technology*, https://www.tvisioninsights.com/our-technology, Complaint Ex. O (screenshot below); Sidhu Interview, https://www.youtube.com/watch?v=xnFypL2JXPE at 3:10-3:20, 4:47-5:06, and 5:32-5:47 ("It is key to us to understand the pose information . . . of the audience. . . . We use deep neural nets to get pose information. We detect humans. We track humans. We kind of have trained our own model for tracking. We use something called graph cut algorithms to track the path of people across multiple frames.").) 
analyzing, using an audience detector, the first image to attempt to	The TVision computer uses the image captured by the webcam to detect a user's head appearing in the image. (*See TVision, Our Technology*, https://www.tvisioninsights.com/our-technology, Complaint Ex. O (screenshot below); Sidhu Interview, https://www.youtube.com/watch?v=xnFypL2JXPE at 3:10-3:20, 4:47-5:06, and 5:32-5:47 ("[i]t is key to us to understand the pose information . . . of the audience. . . . We use deep neural nets to get pose information. We detect humans. We track humans. We kind of have trained our own model for tracking. We use something called graph cut algorithms to track the path of people across multiple frames.).)

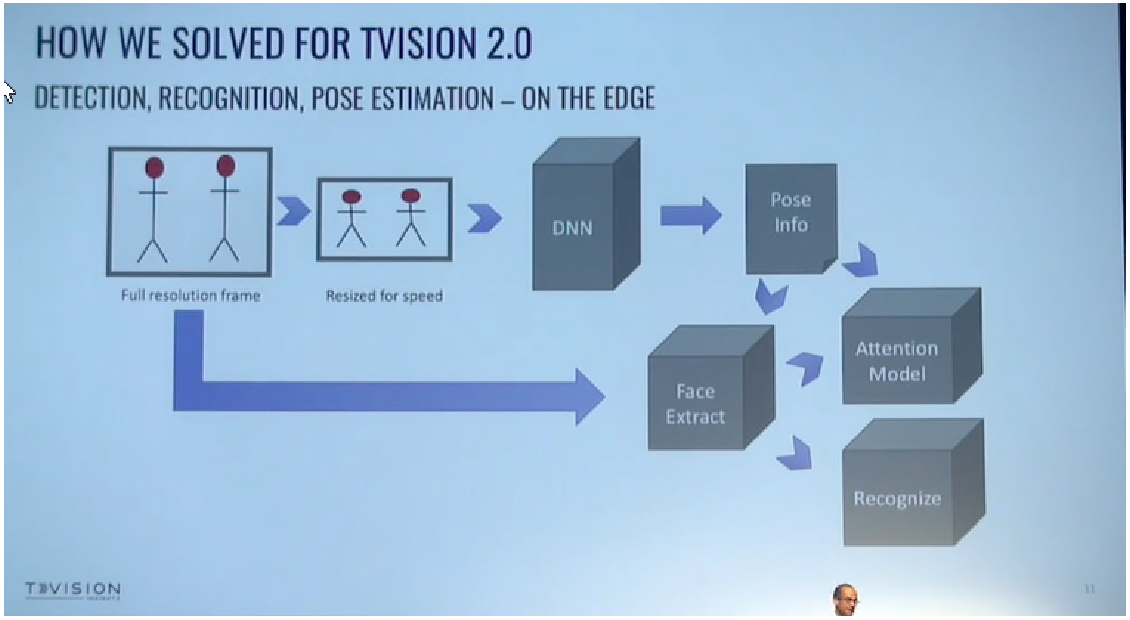
Element of Claim 16	TVision Insights, Inc.
detect a head in the first image;	 Source: https://www.tvisioninsights.com/our-technology (last accessed July 2022).
determining, using the audience detector, an orientation of the head with respect to the camera;	The TVision computer uses computer-vision technology to determine a head pose of the detected head. (*See* Sidhu Interview, https://www.youtube.com/watch?v=xnFypL2JXPE at 5:32-5:47 ("But eventually, all of the pose data makes its way to our attention measurement. And that's where we extract out the head pose, the facial features, and all that to compute whether or not somebody is paying attention and what their gaze is.") (screenshot below).)

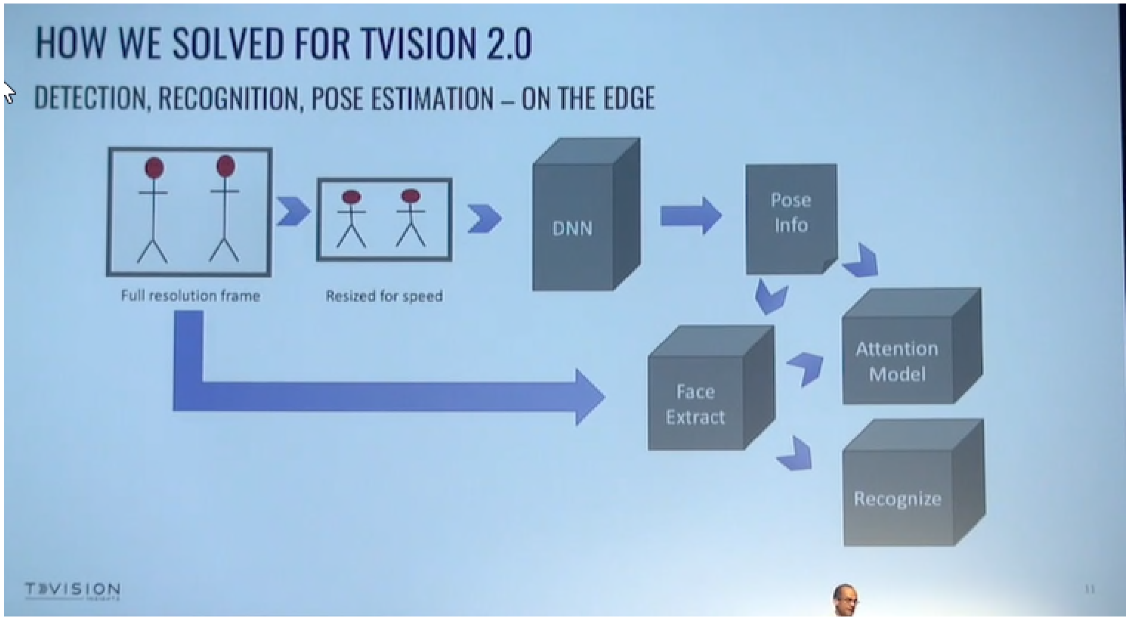
Element of Claim 16	TVision Insights, Inc.
	
determining, using the audience detector, audience identification information based on an indication that the head matches a known person associated with	The TVision computer applies facial recognition to images captured by the webcam to detect who is watching the TV. (*See TVision, Our Technology*, https://www.tvisioninsights.com/our-technology, Complaint Ex. O; *Mytvpanel, About Us*, https://www.mytvpanel.com/about, Complaint Ex. P ("The system starts out in Training Mode, where it captures headshot images of your household members' faces from forehead-to-chin, ear-to-ear. . . . . . The system uses those images to tell the viewers apart and map in their demographics from your household profile."); Sidhu Interview, https://www.youtube.com/watch?v=xnFypL2JXPE at 5:32 (depicting recognition based on facial features) (screenshot below).)

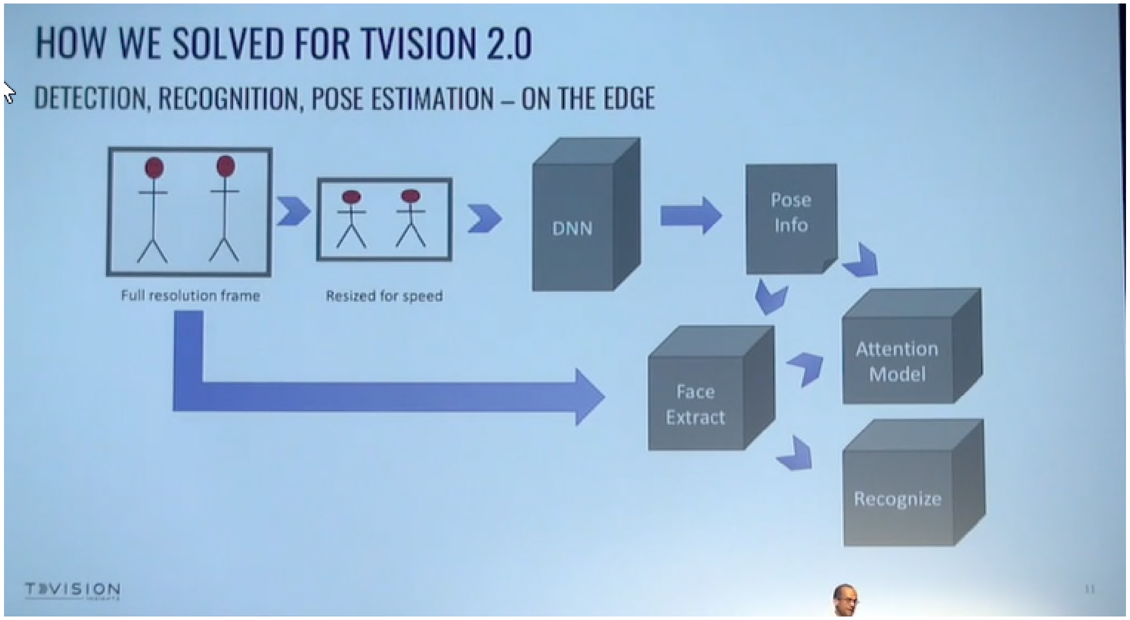
Element of Claim 16	TVision Insights, Inc.
the audience detector; and	
providing the content identifying data and the audience identification information to a data collection facility.	The TVision computer is connected to a household's internet network using an Ethernet cable or WiFi. (*See Mytvpanel, About Us*, https://www.mytvpanel.com/about, Complaint Ex. P.) The TVision computer transmits attention data and the matching program or commercial name to a cloud database using network interface circuitry within the TVision computer. (TVision, "Join the TVision Panel" https://vimeo.com/295447727/3506f24b2b?embedded=true&source=video_title&owner=90679125 at 1:32 ("After completing training mode, the system can tell which person is in the room and when their eyes are on the screen. This attention data is transmitted back to TVision in a text file."); *Mytvpanel, About Us*, https://www.mytvpanel.com/about, Complaint Ex. P (stating that the matching program or commercial name is sent back to TVision).)

Element of Claim 18	**TVision Insights, Inc.**
18. The method of claim 16, further comprising: reducing a resolution of the first image of the media exposure environment to obtain a reduced-resolution image, and	The TVision computer reduces the resolution of images captured by the webcam. (*See* Sidhu Interview, https://www.youtube.com/watch?v=xnFypL2JXPE at 5:20-5:24 ("We have to shrink the frame before we pass it to the neural net.") (screenshot below).) 

<table>
<tr><th>Element of Claim 18</th><th>TVision Insights, Inc.</th></tr>
<tr><td>wherein the determining of the orientation of the head includes determining the orientation of the head with respect to the camera using the reduced-resolution image.</td><td>The TVision computer provides the resized frame to the deep neural network (“DNN”), which determines the pose information using the reduced-resolution image. The TVision computer uses the pose information as the basis to extract the head pose information. (See Sidhu Interview, https://www.youtube.com/watch?v=xnFypL2JXPE at 5:32-5:47 (“But eventually, all of the pose data makes its way to our attention measurement. And that’s where we extract out the head pose, the facial features, and all that to compute whether or not somebody is paying attention and what their gaze is.”) (screenshot below).)

</td></tr>
</table>

Element of Claim 19	**TVision Insights, Inc.**
19. The method of claim 18, further comprising: identifying a region corresponding to the head within the first image from which the reduced-resolution image was obtained;	As explained above, the pose information output by the DNN is determined using a resized (reduced-resolution) frame. The TVision computer extracts facial signatures from a region of a full-resolution frame using the pose information output by the DNN. (*See* Sidhu Interview, https://www.youtube.com/watch?v=xnFypL2JXPE at 5:32-5:47 ("But eventually, all of the pose data makes its way to our attention measurement. And that's where we extract out the head pose, the facial features, and all that to compute whether or not somebody is paying attention and what their gaze is.") (screenshot below).) 

Element of Claim 19	TVision Insights, Inc.
analyzing a portion of the second image corresponding to the identified region; and	As explained above, the pose information output by the DNN is determined using a resized (reduced-resolution) frame. The TVision computer extracts facial signatures from a region of a full-resolution frame using the pose information output by the DNN. (*See* Sidhu Interview, https://www.youtube.com/watch?v=xnFypL2JXPE at 5:32-5:47 ("But eventually, all of the pose data makes its way to our attention measurement. And that's where we extract out the head pose, the facial features, and all that to compute whether or not somebody is paying attention and what their gaze is.") (screenshot below).) 
based on the analyzing of the region of the second image, determining whether the head matches the known person.	The TVision computer applies facial recognition to images captured by the webcam to detect who is watching the TV. (*See TVision*, *Our Technology*, https://www.tvisioninsights.com/our-technology, Complaint Ex. O; *Mytvpanel*, *About Us*, https://www.mytvpanel.com/about, Complaint Ex. P ("The system starts out in Training Mode, where it captures headshot images of your household members' faces from forehead-to-chin, ear-to-ear. . . . . . The system uses those images to tell the viewers apart and map in their demographics from your household profile."); Sidhu Interview, https://www.youtube.com/watch?v=xnFypL2JXPE at 5:32 (depicting recognition based on facial features) (screenshot below).)

Element of Claim 19	TVision Insights, Inc.
	

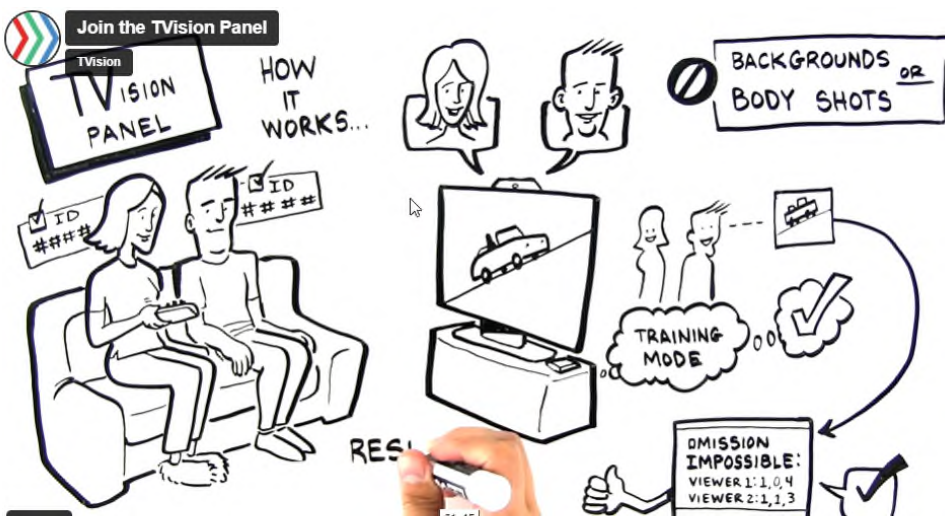
Element of Claim 20	TVision Insights, Inc.
20. The method of claim 19, wherein the audience identification information includes an identifier associated with the known person, and further including recording the identifier associated with the known person in a memory.	The TVision computer generates records of which household members watched which content. The records include anonymized identifiers for each household member that are stored in memory. (TVision, "Join the TVision Panel" https://vimeo.com/295447727/3506f24b2b?embedded=true&source=video_title&owner=90679125 at 1:25-1:42 ("The system uses those [headshot] images to tell the viewers apart and assign an anonymized ID for each person. After completing training mode, the system can tell which person is in the room and when their eyes are on the screen. This attention data is transmitted back to TVision in a text file.") (screenshot below).) 

EXHIBIT R

SHAZAM Search for music

"Our mission is to help people recognize and engage with the world around them"

Shazam is a mobile app that recognises music around you. It is the best way to discover, explore and share the music you love. Shazam connects more than **1 billion** people. It took us 10 years to reach 1 billion Shazams and now we deliver 1 billion song results every month! It's an amazing app, available now in the Apple and Android stores. And we're always looking for new and innovative ways to delight our users.

HOW DID WE GET HERE?

2002

Dialling “2580” on your phone and holding it up to the music was all you had to do to use Shazam. Users were sent an SMS message telling them the song title and the name of the artist.

2008

Shazam was one of the first apps in the brand new Apple App Store.

2011

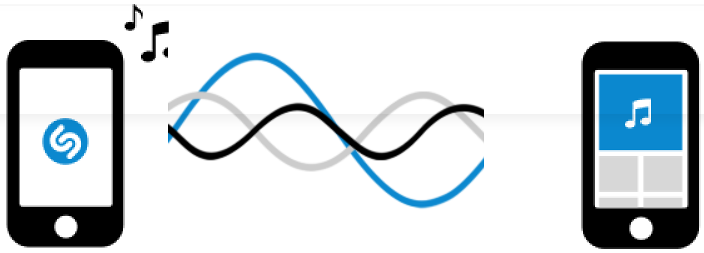
In addition to music, the app was extended to let you Shazam TV programs and ads to get special offers and more information on what you were watching.

TODAY

Shazam joins the Apple family and continues to improve the music discovery journey for all of its users.

HOW DOES SHAZAM WORK?

This is a question we get often asked. Here is a quick summary of the three main steps involved from the moment you Shazam until the magic happens.



Let's say you are in a shop and you like the music you're hearing. Start the app and tap the Shazam button.

A digital fingerprint of the audio is created and, within seconds, matched against Shazam's database of millions of tracks.

You are then given the name of the track and the artist and information such as lyrics, video, artist biography, concert tickets and recommended tracks.

SHAZAM

English

Company

About Us

Careers

Help for Apple Devices

Help for Android Devices

ShazamKit for Developers

Apple Music Offer

Legal

Terms

Privacy

Manage Your Data

GET IT ON Google Play

Download on the App Store

Available in the Chrome Web Store

Available on Galaxy Store

Follow Us

EXHIBIT S

ACRCloud

Home Solutions Industries Login English Sign Up

Live Channel Detection

Detecting Live Channels And Time-Shifting Channels In Scale

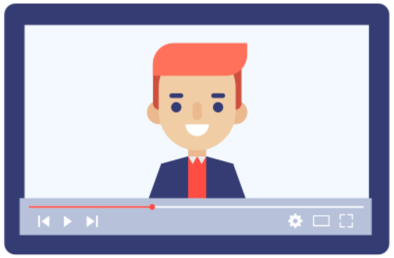
Trusted by

KÖRK Alibaba Group NBA TVISION INSIGHTS ecglobal KÖRK

Detecting Live Channel

Find Out Which Channel Audiences Are Watching Or Listening To.

Detecting Time-Shifting Content



ACRCloud stores the time-shifting content automatically, so when the audiences are watching the time-shifting content, you know exactly what they have watched.

How It Works

ACRCloud collects the live feed from TV or radio stations in real-time and enables the channel to be detected at the exact point of broadcast from any of the user's mobile devices. With live content-generated audio fingerprints, supplementary information and interactive campaigns can be also triggered to appear on the viewer's second screen devices.

Technology In This Solution

Audio Fingerprinting

Learn More

Industries Who Use This Solution

Broadcasters

TV & Radio Stations

Learn More →

Advertising & Big Data

Big Data Platforms, Advertising Agencies

Learn More →

Contact Press Blog Docs Brand Guidelines Terms of Use Privacy Policy

Music Submission 京ICP备15035697号-1

EXHIBIT T

ACRCloud

Home Solutions Industries Login English Sign Up

Advertising & Big Data

Big Data Platforms & Advertising Agencies

Customer Stories

RealityMine has integrated ACRCloud's audio content recognition SDK in its passive metering app to identify specific commercials. The data is analyzed along with personal location behavior to draw insights about consumer behavior from commercial viewings and real-world store visits.

We were attracted by ACRCloud's range of up-to-date SDKs, developer tools and high scalability. Integration of the Android SDK was straightforward and when an opportunity to enhance the performance on low-end devices was discovered, an improvement was delivered very rapidly.

Graham Dean, CTO of RealityMine

Got any questions? I'm happy to help!

"

We trust ACRCloud when it comes to fingerprinting, the way they have grown over the past few years is impressive!

David Weiszfeld, CEO of Soundcharts

Soundcharts works with ACRCloud to monitor music airplay on radio stations around the world. Soundcharts integrates with ACRCloud's API of broadcast monitoring to receive all music airplay data for each station. It then distributes monitoring results to artists or labels who are using Soundcharts.

Soundcharts

Customers With Similar Scenarios

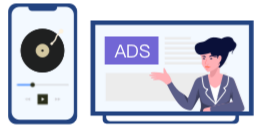
TVISION INSIGHTS yacast KÖRK ecglobal TVISION INSIGHTS

Start Integrating ACRCloud With Your Project

Sign Up

User Scenarios

Got any questions? I'm happy to help!



Track Ads Or Music Usage

Get to know when and where the Ads or music have been played.

Audience Measurement On TV

Get insights of the TV audiences by measuring when and what content they have watched.

Related Solutions

Broadcast Monitoring

The API-First Broadcast Monitoring Services For Your Business

Learn More →

Audience Measurement

More Accurate & Convincing

Learn More →

Offline Recognition

Detecting Custom Content On Devices Without Network Connection

Learn More →

Contact Press Blog Docs Brand Guidelines Terms of Use Privacy Policy

Music Submission 京ICP备15035697号-1

Got any questions? I'm happy to help!